Investigation into Information Capacity of Nonlinear Optical Fibre Communication Systems

Mykyta Shevchenko

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
of
University College London.

Department of Electronic & Electrical Engineering
University College London

March 13, 2019

I, Mykyta Shevchenko, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

The optical fibre is a ubiquitous transmission medium since it is able to provide both high speed and low loss. Optical fibre transmission systems carry 99 % of the world's telecommunication traffic. The emergence of new services and Internet applications gives rise to the exponentially increasing demand for higher transmission data rates, motivating the search for new methods to enhance the capacity of optical fibre systems. However, due to the presence of power-dependent signal degradation effects (the optical Kerr effects) together with bandwidth limitations constrained by the low-loss region of the fibre, the current optical fibre communication infrastructure is unable to cope with the ever-growing demand for data rates. The capacity of an optical fibre channel remains unknown and is an open research question.

The PhD research described in this thesis aimed to theoretically investigate the capacity of the nonlinear optical fibre channel using information-theoretic tools with the view to improving information data rates of optical fibre networks. The first part of the thesis is concerned with a comprehensive study of Kerr nonlinearity-compensated dispersion-unmanaged ultra-wide bandwidth optical fibre communication systems. The bounds on information rate, based on the proposed model, which takes into account the fundamental limitations due to nonlinear interactions between optical signal and amplifier noise, were accurately estimated. The second part deals with the application of the so-called integrability property (the general ideas based around nonlinear Fourier transform (NFT)) of a lossless and noiseless nonlinear Schrödinger equation (NLSE). A new *non-Gaussian* channel model for soliton-based transmission, in which data is assumed to be embedded into the imaginary part of the nonlinear discrete spectrum was proposed for the first time. New asymptotic semi-analytic approximations for non-decaying capacity bounds have been derived. The theoretical results of this research can be considered as an important first step towards the ultimate capacity limits of nonlinear optical communication links.

Impact Statement

It is almost impossible to imagine a modern society without Internet networking technologies, which are able to provide a fast, reliable and secure exchange of information. There is still a common belief that the Internet data are somehow transmitted by means of satellites; however, this opinion does not correspond to the reality. In fact, optical fibre communication systems are currently responsible for the overwhelming majority of global information traffic. These systems have already experienced a relatively long process of improving their performance from speeds of few gigabits per second in the 1990s to astonishing hundreds of terabit per second today mainly by means of increasing their engineering complexity and sophistication.

The critical physical phenomenon distinguishing optical fibre communication systems from either satellite or wireless transmission systems is the presence of the optical Kerr nonlinearity (*i.e.,* the intensity-dependent refractive index) in a fibre. This effect has a detrimental impact on the information capacity (that is, on the maximum error-free transmission information rate), and ultimately creates formidable obstacles to fulfil ever-increasing optical fibre network capacity demands. Notably, the theoretical analysis of current optical fibre communication systems keeps exploiting well-established methods of information theory, which are mostly suitable for linear wireless communication systems, and hence, cannot be applied to inherently nonlinear communication systems, such as the optical fibre transmission systems. In addition, the ongoing rapid growth of world traffic demand is restlessly pushing towards a looming information capacity limit, which is imposed by the Kerr nonlinearity-induced distortions.

In this thesis, theoretical investigations into the information capacity of nonlinear optical fibre communication systems were carried out, which have great potential to be useful and practically implementable in real advanced optical fibre communication systems. In particular, it was shown that increasing the transmission bandwidth could significantly increase system throughput, and these increases were quantified. However, it was also shown that

the nonlinear distortions due to the Kerr-induced interactions between signal and optical amplifier noise asymptotically reduce the performance of conventional optical fibre systems. These results clearly indicate the need to identify radically new approaches for the encoding, transmission, and processing of information, which can account for the nonlinear nature of the optical fibre.

A new communication system where signals would be encoded in such a way so as to propagate in a linear manner was proposed and analysed. This consists of a train of noise-perturbed optical solitons addressing the modulation of the discrete eigenvalues associated with the so-called nonlinear Fourier transform. A perturbative approximation of the channel law (*i.e.,* the discrete-time input-output relationship) was obtained in closed-form. It was shown that there exist lower bounds on the capacity of inherently nonlinear optical fibre channel that monotonically increase with the effective signal-to-noise ratio, similarly to the linear Gaussian channel. The results are key to overcoming the "capacity crunch" by means of considering a significantly more challenging problem of the capacity in bit per second per unit bandwidth.

Acknowledgements

First and foremost, I wish to express my profound gratitude to my principal supervisor Prof. Polina Bayvel for her continuous support throughout my PhD research, as well as for her professional guidance and great magnanimity. I also offer my sincerest thanks to my second supervisor Dr. Robert Killey for his kind help and loyalty. There is no doubt that without their professional advice and expertise, the thesis would never have been possible.

Besides my supervisors, I am highly obliged to all my fellow students in the Optical Networks Group (ONG) in the 808 office at UCL. I would also like to thank all my collaborators. Especially to Dr. Tianhua Xu and Dr. Domaniç Lavery for teaching me the theory of digital signal processing (DSP) as well as nonlineaity compensation algorithms, to Dr. Jaroslaw Prilepsky and Prof. Sergei Turitsyn who introduced me to physics of optical solitons and the concept on the nonlinear Fourier transform (NFT), to Mr. Daniel Semrau and Dr. David Ives for deep discussions of the so-called Gaussian noise model of nonlinear "noise-like" distortions, to Dr. Gabriele Liga for explaining me some basic principles of coding theory, to Dr. Alex Alvarado and Prof. Erik Agrell who kindly introduced me to the rigorous theory of information and digital communication. I also owe my special gratitude to Dr. Stanislav Derevyanko for his immense and comprehensive knowledge, as well as his tremendous help in overcoming some mathematical obstacles.

In addition, I would like to take this opportunity to express appreciation to my first physics teacher Anatoly Braga. I would also like to give special thanks to Prof. Igor Lyubchanskii for helping me to make the first steps into the world of science.

Above all, my deepest gratitude goes to my parents: my mother Natalia, my father Alexander, my sister Anna, and my niece Anastasia, without their constant support this thesis would never have turned into reality.

Finally, I gratefully acknowledge a financial support through a UCL Graduate Research Scholarship (GRS) and the UK EPSRC UNLOC project.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 A brief history of optical communications

The nature of light propagation in optical fibres relies on the phenomenon known as total internal reflection. In 1841, Swiss physicist Jean-Daniel Colladon first demonstrated that, using a water-air interface, it is possible to guide light with a falling stream of water. This experiment now represents one of the key principles of modern optical fibre. In 1854, at the Royal Institution, Irish physicist John Tyndall repeated Colladon's experiment of the transmission of light along a water stream arising from a hole in the side of a tank, introducing the idea of guided light to follow a specific path. The first glass fibres were fabricated in the 1920s for medical applications. However, the use of such fibres for the purposes of optical communications was entirely infeasible, owing to their extremely high attenuation ($\sim 10^3\,\mathrm{dB\,km^{-1}}$). The year 1962 can be confidently considered as a starting point in modern optical communications, when electrically-pumped semiconductor lasers were first reported independently by several research groups [3, 4][1]. In 1966, for the first time, the dielectric single-mode silica-based fibre was proposed as a waveguide for communications at optical frequencies by Kao and Hockham, and the limitations due to attenuation were discussed [6]. It was demonstrated that optical fibres could provide transmission with a relatively low attenuation $\sim 20\,\mathrm{dB\,km^{-1}}$, determined by practical system considerations at the time, at optical carrier frequency of ($\nu_0 \approx 193.4\,\mathrm{THz}$). The attenuation of the state-of-the-art single mode optical fibres has since been greatly reduced by further purification of fused silica that led to the fibre loss to approximately $0.2\,\mathrm{dB\,km^{-1}}$ around the $\lambda_0 = 1550\,\mathrm{nm}$ spectral region [7].[2] Notably, in comparison with microwaves,

[1] However, the first semiconductor laser operating at room temperature was reported eight years later in 1970 [5].

[2] A fibre with the lowest currently recorded attenuation was recently reported by [8, 9].

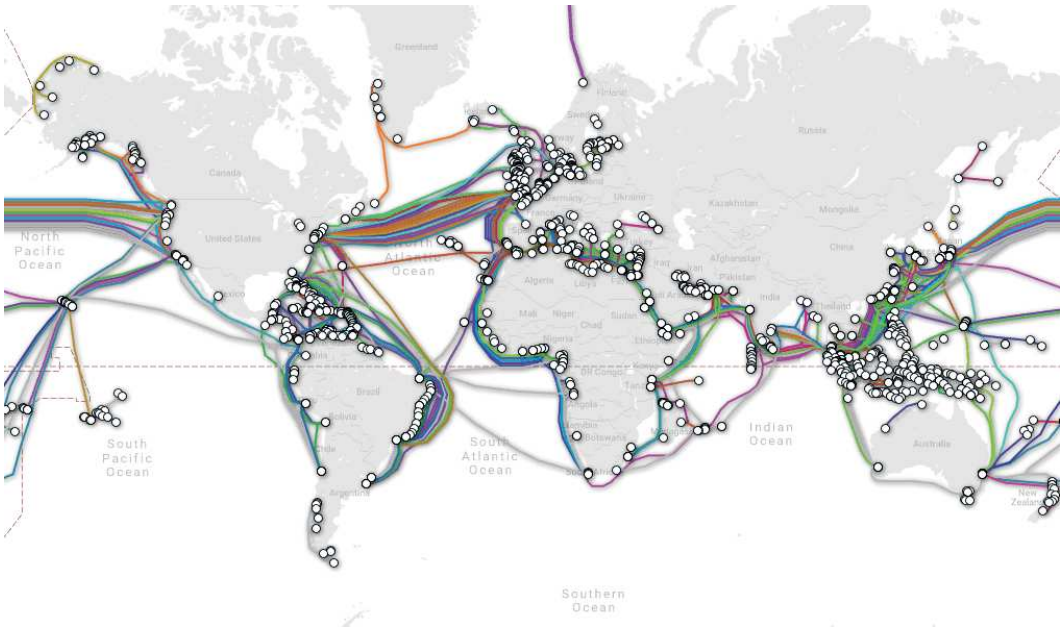

Figure 1.1: Submarine optical fibre infrastructure underpins telecommunications and the Internet, provided and managed by TeleGeography [1].

which enable wireless communications, the frequencies of light that are carried in fibres are roughly four orders of magnitude higher, therefore, the possible modulation transmission bandwidth increases accordingly. It is also worth mentioning that the optical fibre remains unsurpassed, with no other known medium, which would be able to underpin the immense demands for information data rate, reliability as well as energy efficiency.

The extremely large bandwidth (up to 400 nm), determined by the low loss region of the fused silica, between 1300–1700 nm, makes fibre an excellent medium in terms of transmitting information signals with no required optical signal amplification over fairly long distances (up to 100 km). However, even with all these advantages a long-haul optical communication system still requires periodic signal amplification, which makes optical amplification via the invention of the erbium-doped fiber amplifier (EDFA) in the late 1980s hugely important [10, 11, 12]. This amplification approach completely displaced much more expensive optoelectronic repeaters, which were used to detect optical signals at short distances, and then re-transmit them at higher power. Indeed, with the advent and subsequent commercialisation of ultra-wideband optical amplifiers, such as above-mentioned EDFA as well as distributed Raman amplifier (DRA)s [13, 14], it became possible to extend the usable fibre bandwidth up to approximately ∼35 nm for EDFA (known as the conventional or C-band), and up to ∼100 nm for the distributed Raman amplifier or combined EDFA

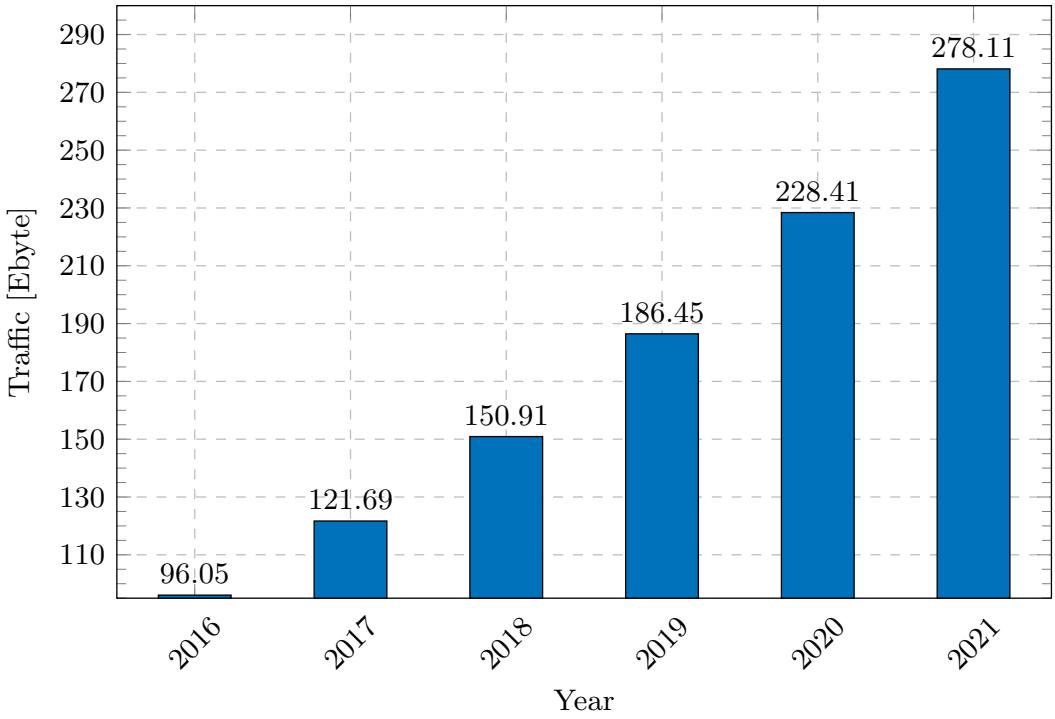

Figure 1.2: The past and predicted increase in the total Internet traffic, according to Cisco Systems, Inc. [2].

(operating over the C- and L-(long) bands), which in conjunction with WDM enabled the rise of enable multi-user optical communications, and significantly improved the overall data rates (up to tens of terabits per second).

1.2 The capacity "crunch" problem

Currently, optical fibre communications have become ubiquitous in modern society, since it made it possible to exchange information fast and reliably over great distances. It has been estimated that about 99 % of the Internet data traffic are carried by means of optical fibres (see, *e.g.,* [15]). The optical fibre-based networks currently underpin the main part of the national and international communication infrastructure. Fig. 1.1 clearly demonstrates the significance of optical fibre for transmitting a tremendous amount of information data across the continents via established submarine cables. The overwhelming bulk of the world's optical fibre telecommunication systems carrying international traffic have undergone a long process of increasing engineering complexity and sophistication. For instance, the successful implementation of the "fifth generation" optical transceivers and networks operating with coherent optical detection has enabled multilevel signal modulation formats, and advanced digital signal processing techniques, leading to the possibility of chan-

nel transmission rates exceeding 100 Gbps [16]. The key to this breakthrough was the mitigation of linear transmission impairments, such as chromatic and polarisation mode dispersion. Nonetheless, the key *physical effects* affecting the performance of optical fibre communication systems remain largely the same. These are: chromatic dispersion (*i.e.,* second-order dispersion), fibre nonlinearity[3], which occurs due to the intensity-dependent refractive index (*i.e.,* third-order optical nonlinear effect), and the additive optical noise due to amplified spontaneous emission (ASE) arising from the optical amplification process. Furthermore, the recent advent of bandwidth-hungry applications, such as high-definition video streaming, IP-TV, online real-time gaming, peer-to-peer file sharing, *etc.* has quickly spurred demand for higher transmission rates. The long-range forecast of demand is depicted in Fig. 1.2, which indicates a tremendous increase in the global Internet traffic, according to Cisco's statistics and predictions.

It is now widely accepted that to accommodate the anticipated growth of data rates in optical fibre telecommunication systems (see, Fig. 1.2), the efficiency of the available optical fibre bandwidth use (*i.e.,* the spectral efficiency, or the overall capacity) needs to be improved [18]. However, steadily pushed by the rapidly growing traffic demands, the performance of current coherent multi-channel optical fibre communication systems is now progressively approaching an impending limit on the data rate of single-mode fibre, which is mainly set by the fibre Kerr nonlinearity [19]. In contrast to a free-space wireless communication channels (*i.e.,* the linear channels), the optical fibre communication channel is inherently nonlinear, and thus, the spectral efficiencies usually exhibit a peaky behaviour and decay at high launched power regime; this is often referred to as the "nonlinear Shannon limit", *i.e.,* the apparent "limit", which is imposed by the intensity-dependent refractive index, also known as the optical Kerr-induced nonlinearity limit. Such a limit has been suggested to be a major constraint to the information carrying optical fibre capacity [19, 20]. It is worth emphasising that using the term "Shannon limit" in the context of a NL channel is not entirely appropriate, since the existence of such a "limit" has not yet been rigorously proven. Indeed, the Shannon capacity has been defined under the assumption of a Gaussian channel law, whilst the statistics of the nonlinear optical fibre channel in the high launched power regime is no longer Gaussian, and the exact channel model has yet to be found (see, *e.g.,* [21, Sec. 10] and refs therein).

[3]Amongst the first who considered the Kerr nonlinearity in an optical fibre were R. H. Stolen and A. Ashkin [17].

Although the bandwidth of optical fibre transmission systems is large (*e.g.,* up to 50 THz at 1550 nm operating wavelength), optical fibre communication systems are ultimately bandwidth-limited. The bandwidth limitation combined with difficulty in overcoming fibre nonlinearity is expected to result in a so-called infamous *"capacity crunch"* problem, that is, the inability of the current optical fibre network infrastructure to cope with the ever-growing capacity demand, which caps the rate increase of error-free data transmission. The term "capacity crunch" was applied to communication networks and was first highlighted by Chraplyvy in 2009 [22], which aroused interest in developing new approaches to system design, including the use of, for instance, closer channel spacing, denser signal modulation formats, enabled by coherent detection. Despite the significant progress, many aspects related to the looming "capacity crunch" problem remain as open research questions [23, Sec. 1], [24, 25]. The research, described in this thesis, addressed two key challenges. First, an estimation of achievable information rates for ultra-wide bandwidth optical fibre communication systems considering limitations due to signal-noise interactions. Also, a new channel model for nonlineal optical fibre channel are proposed. The lower bounds on channel capacity are analytically estimated. Even though the true channel capacity, *i.e.,* the maximum achievable data rate for the nonlinear optical channel is unknown exactly, the analytical bounds and capacity estimates derived in this thesis can give important insights into future optical communication system designs.

1.3 Thesis structure

The remainder of the thesis is organised as follows.

Chapter 2 introduces a brief description of the fundamental theoretical tools needed to understand the mathematical modelling underlying optical fibre communications. First of all, the chapter gives concise information-theoretic background required to analyse an information channel capacity. Secondly, this chapter highlights the physical phenomena that underlie the linear and nonlinear propagation of an arbitrary optical waveforms in an optical fibre. The most widespread phenomenological analytical and efficient numerical approaches, which describe a continuous-time optical fibre communication channel are highlighted. A discrete-time optical fibre communication system is also described.

Chapter 3 first makes an overview of analytical modelling of the most used optical amplification schemes, using lumped erbium-doped fibre amplifiers and distributed (both ideal and non-ideal) Raman amplification. Then,

the conventional Gaussian noise models to assess the distortions due to the presence of optical Kerr effect in a fibre are introduced and comprehensively reviewed. The final part of this chapter is dedicated to a signal-noise interaction process in nonlinearity-compensated optical links. The model was modified by including the second-order contributions of signal-noise interaction together with modulation-dependent nonlinear effects. The performance of multi-channel nonlinearity-compensated optical fibre communication systems are estimated via proposed analytical approach and numerical simulations.

In **Chapter 4**, fundamental Kerr nonlinearity limits on achievable information rates of ultra-wide bandwidth Nyquist-spaced WDM optical fibre communication systems are analytically evaluated within the framework of first-order perturbation analysis. To approach the estimated limits, advanced modulation formats in conjunction with probabilistically-shaped signal constellation were considered. In addition, the capacity gains, which might be potentially achieved by increasing optical signal bandwidths uo to 400 nm are theoretically analysed.

Chapter 5 is concerned with a new optical fibre communication scheme in which information is embedded into a train of well-separated (in time) solitons amplitudes, and then effectively detected through the so-called forward nonlinear Fourier transform. A perturbation-based discrete-time non-Gaussian channel model is developed, and then used to analytically study lower bounds on the channel capacity under an average power constraint. Some practically relevant cases, such as amplitude shift-keying constellations, to approach derived analytical capacity lower bounds, are also numerically analysed.

In **Chapter 6**, general conclusions on the work carried out are described, and the questions left for further investigation are highlighted.

Finally, the **Appendix** supplements the results of Chapter 5 adding detailed proofs of the theorems. The brief description of the used analytical and numerical methods was also complemented.

1.4 Thesis key contributions

- The influence of second-order signal-noise interactions on the performance of nonlinearity-compensated optical fibre communication systems in the presence of modulation-format-dependent nonlinearity was investigated. This comprehensive study (Chapter 3) was published in [26].

- A quasi-Monte Carlo approach to theoretically predict the performance of wideband Nyquist-spaced optical communication systems using 40 nm

EDFA and 100 nm DRA was effectively implemented. The impact of nonlinear compensation together with a signal probabilistic shaping was analysed (Chapter 4). Some of the results are in [27], and [28].

- A comprehensive study of the impact of chromatic dispersion slope of the performance metrics of Nyquist-spaced ultra-wide bandwidth optical communication systems (Sec. 4.2).

- The fundamental upper bounds on the capacity, imposed by the optical Kerr effect in a fibre, are analytically assessed in Sec. 4.3. To assess to what extent the achievable capacity of the nonlinear optical fibre channel can be increased by means of a complete nonlinearity compensation. The study was included in [29].

- A new generalised discrete-time non-Gaussian channel model based on the perturbative conditional PDF for soliton amplitudes was developed (Chapter 5). The exact and approximate capacity lower bounds in bits per channel use, which show an unbounded growth with the effective SNR similar to the linear Gaussian channel, were analytically derived. The thorough research of the proposed channel model with applications to practically relevant optical soliton amplitude modulation was published in [30, 31].

1.5 List of publications

The research work presented in this thesis has also lead to the following publications:

Journal papers

1. **N. A. Shevchenko**, S. A. Derevyanko, J. E. Prilepsky, A. Alvarado, P. Bayvel, and S. K. Turitsyn, "A lower bound on the capacity of the non-central chi channel with applications to soliton amplitude modulation," IEEE Transactions on Communications, vol. 66, no. 7, pp. 2978–2993, 2018.

2. **N. A. Shevchenko**, T. Xu, D. Lavery, G. Liga, D. J. Ives, R. I. Killey, P. Bayvel, "Modeling of nonlinearity-compensated optical communication systems considering second-order signal-noise interactions," Optics Letters, vol. 42, no. 17, pp. 3351–3354, 2017.

3. T. Xu, B. Karanov, **N. A. Shevchenko**, D. Lavery, G. Liga, R. I. Killey, P. Bayvel, "Digital nonlinearity compensation in high-capacity optical

communication systems considering signal spectral broadening effect," Nature Scientific Reports, vol. 7, no. 1, 12986, 2017.

4. T. Xu, **N. A. Shevchenko**, D. Lavery, D. Semrau, G. Liga, A. Alvarado, R. I. Killey, P. Bayvel, "Modulation format dependence of digital nonlinearity compensation performance in optical fibre communication systems," Optics Express, vol. 25, no. 4, pp. 3311–3326, 2017.

5. D. Semrau, T. Xu, **N. A. Shevchenko**, M. Paskov, A. Alvarado, R. I. Killey, and P. Bayvel, "Achievable information rates estimates in optically amplified transmission systems using nonlinearity compensation and probabilistic shaping," Optics Letters, vol. 42, no. 1, pp. 121–124, 2017.

6. B. Karanov, T. Xu, **N. A. Shevchenko**, D. Lavery, R. I. Killey, and P. Bayvel, "Span length and information rate optimisation in optical transmission systems using single-channel digital backpropagation," Optics Express, vol. 25, no. 21, pp. 25353–25362, 2017.

7. P. Bayvel, R. Maher, T. Xu, G. Liga, **N. A. Shevchenko**, D. Lavery, A. Alvarado, R. I. Killey, "Maximizing the optical network capacity," Philosophical Transactions of the Royal Society A, vol. 374 no. 2062, pp. 20140440, 2016.

Conference papers

8. T. Xu, B. Karanov, **N. A. Shevchenko**, D. Lavery, G. Liga, Z. Li, D. Jia, L. Li, L. Kanthan, R. I. Killey, P. Bayvel, "Spectral broadening effects in optical communication networks: Impact and security issue," IEEE International Conference on Advanced Infocomm Technology (ICAIT), Stockholm, Sweden, 2018.

9. B. Karanov, T. Xu, **N. A. Shevchenko**, D. Lavery, G. Liga, R. I. Killey, P. Bayvel, "Digital nonlinearity compensation considering signal spectral broadening effects in dispersion-managed systems," Optical Fiber Communications Conference and Exposition (OFC), San Diego, California, USA, 2018.

10. E. Sillekens, D. Semrau, G. Liga, **N. A. Shevchenko**, Z. Li, A. Alvarado, P. Bayvel, R. I. Killey, D. Lavery, "A simple nonlinearity-tailored

probabilistic shaping distribution for square QAM," Optical Fiber Communications Conference and Exposition (OFC), San Diego, California, USA, 2018.

11. T. Xu, **N. A. Shevchenko**, B. Karanov, D. Lavery, L. Galdino, A. Alvarado, R. I. Killey, P. Bayvel, "Nonlinearity compensation and information rates in fully-loaded C-band optical fibre transmission systems," European Conference on Optical Communication (ECOC), Gothenburg, Sweden, 2017.

12. T. Xu, **N. A. Shevchenko**, B. Karanov, G. Liga, D. Lavery, R. I. Killey, P. Bayvel, "Digital nonlinearity compensation in high-capacity optical fibre communication systems: Performance and optimisation," Advances in Wireless and Optical Communications (RTUWO), Riga, Latvia, 2017.

13. **N. A. Shevchenko**, T. Xu, D. Semrau, G. Saavedra, G. Liga, M. Paskov, L. Galdino, A. Alvarado, R. I. Killey, P. Bayvel, "Achievable information rates estimation for 100-nm Raman-amplified optical transmission system," European Conference on Optical Communication (ECOC), Düsseldorf, Germany, 2016.

14. **N. A. Shevchenko**, J. E. Prilepsky, S. A. Derevyanko, A. Alvarado, P. Bayvel, and S. K. Turitsyn "A lower bound on the per soliton capacity of the nonlinear optical fibre channel," IEEE Information Theory Workshop, Jeju Island, South Korea, 2015.

Chapter 2

Mathematical tools for optical fibre communication

This chapter presents a brief description of the fundamental mathematical principles, which are essential for the theoretical analysis of fibre-optic communication systems. The basic theoretical concepts described in this chapter will be used throughout the thesis.

2.1 Information-theoretic background

In 1948, American mathematician and engineer Claude Shannon published his pioneering paper, which can be considered as the birth of modern information and communication theory. One of the main purposes of information theory is a precise quantification of fundamental limits (benchmarks) in communication systems reflecting the possibility of both transmission and compression of information. It is worth noting that all modern communication systems are essentially designed based on the information-theoretic benchmarks. The main principles of information theory mainly originate from the Claude Shannon's ingenious idea of defining and measuring information. The information theory also ascertains rigorous mathematical relationships between information capacity, reliability and energy. The following chapter provides a concise information-theoretic background required for further analysis.

2.1.1 Shannon entropy

A comprehensive definition of information measure associated with a random variable (RV) X with possible outcomes $X = x$ (realisations) in the alphabet $\mathcal{X}$ and the probability $p_X(x)$ has been introduced first as the average value (statistical mean) of the quantity $-\log p_X(x)$, which is historically called *Shannon entropy* [32]

$$H_X \triangleq -\mathbb{E}\left[\log p_X(X)\right], \tag{2.1}$$

where $\mathbb{E}[\cdot]$ denotes the expectation operator taken over all RVs appearing in its argument. It can be seen that Shannon entropy defined by Eq. (2.1) formally coincides with the so-called Boltzmann entropy, which is defined, in statistical physics, as a logarithmic measure of a number of micro-states of the complex system. Generally speaking, it is also a physical measure of disorder, ambiguity or uncertainty about the system. For instance, the maximum uncertainty corresponds to the situation, when it is the most difficult to predict the possible outcomes, *i.e.,* the minimum information. In its turn, the Shannon entropy can be treated as a measure of randomness or, in other words, a measure of "self-information" of a source, *i.e.,* the average amount of information per symbol of the source. The base of logarithm in Eq. (2.1) denotes the units of Shannon entropy, *e.g.,* if we chose the natural logarithm, the units of entropy H would be nats [nat/sym], if it was the common logarithm, the entropy would measure in Hartley [Hart/sym]. However, it is more common to use the binary logarithm, which gives Shannon entropy the unit of bits [bit/sym].

Accounting for the definition of the mathematical expectation, the Shannon entropy H_X measured in bit per symbol of a discrete RV X with a probability mass function (PMF) $P_X(x) \triangleq \mathbb{P}[X = x]$, $x \in \mathcal{X}$, and the alphabet $\mathcal{X} \triangleq \text{Supp}\,[P_X(x)]$ is defined as[1]

$$H_X \triangleq -\sum_{x\in\mathcal{X}} P_X(x)\log P_X(x). \tag{2.2}$$

If the elements $x \in \mathcal{X}$ are equally likely, *i.e.,* $P_X(x) = \frac{1}{|\mathcal{X}|}$, $\forall x \in \mathcal{X}$, the entropy H_X is then bounded as follows[2]

$$0 \leq H_X \leq \log|\mathcal{X}|\,. \tag{2.3}$$

The average information, which is contained in the channel input and output is defined by the *joint entropy* associated with the joint distribution $P_{X,Y}(x,y)$ with $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, it reads

$$H_{X,Y} \triangleq -\mathbb{E}\left[\log P_{X,Y}(X,Y)\right], \tag{2.4}$$

$$= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{X,Y}(x,y)\log P_{X,Y}(x,y)\,. \tag{2.5}$$

The average information we learned from the channel output given the information we have from the channel input is conveyed by the conditional

[1] The sum is taken over using the convention that $0\log 0 = 0$.

[2] $H_X = 0$ if the input is certain, *i.e.,* there exists an element occurring with probability 1.

probability $P_{Y|X}(y|x)$ and is defined as

$$H_{Y|X}(y|x) \triangleq -\mathbb{E}\left[\log P_{Y|X}(Y|X)\right], \tag{2.6}$$

$$= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{X,Y}(x,y)\log P_{Y|X}(y|x). \tag{2.7}$$

It is worth noting that when the cumulative distribution $P_X(x) \triangleq \mathbb{P}[X \leq x]$ is a continuous function, then the RV X is said to be continuous as well. The function defined as $p_X(x) \triangleq dP_X(x)/dx$ is referred to as the probability density function (PDF) of RV X if and only if $\int_{-\infty}^{\infty} p_X(x)\,dx = 1$ (known as the normalisation condition). Thus, one can define the differential Shannon entropy h_X of a continuous RV X with PDF $p_X(x)$ as follows

$$h_X \triangleq -\int_{\mathcal{X}} p_X(x)\log p_X(x)\,dx, \tag{2.8}$$

where $\mathcal{X} \triangleq \text{Supp}\,[p_X(x)] = \{x\,|\,p_X(x) > 0\}$ is the support set (region) of the RV X. Unlike the entropy H_X, the differential entropy h_X can also be negative.

The conditional differential entropy is defined accordingly

$$h_{Y|X} \triangleq -\int_{\mathcal{X}}\int_{\mathcal{Y}} p_{X,Y}(x,y)\log p_{Y|X}(y|x)\,dx\,dy, \tag{2.9}$$

where $\mathcal{Y} \triangleq \text{Supp}\,[p_Y(y)]$, and the output PDF $p_Y(y)$ is given by

$$p_Y(y) = \int_{\mathcal{X}} p_{Y|X}(y|x)\,p_X(x)\,dx, \tag{2.10}$$

and the conditional PDF $p_{Y|X}(y|x)$ is commonly referred to as a *channel law*.

It should also be emphasised that using bits per symbol as units of entropy has a cogent reason. If we have $N = 2^q$ equally likely symbols, *i.e.,* the probability of each symbol is $p_i = 1/N = 1/2^q\,, \forall i$, therefore, the Shannon entropy is $H = -\mathbb{E}[\log_2(1/2^q)]$. This explicitly means that the entropy of the self-information of the symbol is precisely equal to the symbol bit length. In other words, the number of digits of a binary sequence (*i.e.,* codeword) defined by the source is exactly the entropy of the source.

2.1.2 Relative entropy and mutual information

The relative entropy (*distance*) between the two PDFs $p_X(x)$ and $p_Y(y)$, also known as the Kullback-Leibler divergence (or *discrimination*) is defined

as [33]

$$D\left(p_X(x)\,\|\,p_Y(x)\right) \triangleq \mathbb{E}\left[\log\frac{p_X(X)}{p_Y(X)}\right]. \tag{2.11}$$

It is important to note that the relative entropy is always nonnegative[3], *i.e.,*

$$D\left(p_X(x)\,\|\,p_Y(x)\right) \geq 0, \tag{2.12}$$

with the equality if and only if $p_X(x) = p_Y(x)$ *almost everywhere* [32].

The common information between two RVs can be quantified via the so-called MI. It can be expressed as a relative entropy between the joint distribution $p_{X,Y}(x,y)$ and the product function $p_X(x)\,p_Y(y)$, therefore, it yields

$$I_{X,Y} \triangleq D\Big(p_{X,Y}(x,y)\,\|\,p_X(x)\,p_Y(y)\Big) \tag{2.13}$$

$$= \mathbb{E}\left[\log\frac{p_{X,Y}(X,Y)}{p_X(X)\,p_Y(X)}\right]. \tag{2.14}$$

The MI is, in essence, defined as the reduction of uncertainty of channel input that we get from the knowledge of channel output and *vice versa*. In other words, the value of MI indicates how much on average the realisation y of RV Y *tells* us about the realisation x of RV X.

The above-mentioned definitions of the MI in nats per channel use (or equivalently, nats per symbol) can be held for continuous RVs as follows

$$I_{X,Y} \triangleq \int_{\mathcal{X}}\int_{\mathcal{Y}} p_{X,Y}(x,y)\log\frac{p_{X,Y}(x,y)}{p_X(x)\,p_Y(y)}\,dx\,dy\,. \tag{2.15}$$

The MI can also be readily decomposed as follows

$$I_{X,Y} = h_X + h_Y - h_{X,Y} \tag{2.16}$$

$$= h_X - h_{X|Y} \tag{2.17}$$

$$= h_Y - h_{Y|X} \tag{2.18}$$

$$= I_{Y,X}\,, \tag{2.19}$$

where the joint differential entropy $h_{X,Y}$ in Eq. (2.16) is appropriately given by

$$h_{X,Y} \triangleq -\int_{\mathcal{X}}\int_{\mathcal{Y}} p_{X,Y}(x,y)\log p_{X,Y}(x,y)\,dx\,dy\,. \tag{2.20}$$

[3]It can be straightforwardly proven by using Jensen's inequality and the concavity of the log function.

The nonnegativity property Eq. (2.12) in conjunction with the definition Eq. (2.13) leads to the nonnegativity of the MI

$$I_{X,Y} \geq 0\,, \tag{2.21}$$

with the equality if and only if X and Y are statistically independent.

2.1.3 Information channel capacity

It is customary for the implementation of information-theoretic analysis to characterise a communication channel via the conditional PDF $p_{Y_k|X_k}(y_k|x_k)$, where $k \in \mathbb{N}$ denotes the discrete time index. Here, the knowledge of the finite-dimensional conditional output distribution function $p_{Y_k|X_k}(y_k|x_k)$ is utterly vital and is sometimes referred to as an information channel model (or a "channel law"). The elements (symbols) of input vector $X_k = [X_1, ..., X_k]$ and output vector $Y_k = [Y_1, ..., Y_k]$ can be either real or complex RVs. According to Shannon theory, the channel capacity can be defined as a maximum rate for a possible reliable communication. In other words, the channel capacity gives the highest information rate for which the sequence of codes with vanishing probability of error can be found (see Shannon's channel *coding theorem*) [34, 32]. For the information *stable* channel [35] with memory (which is the case of nonlinear channel), this statement reflects the general definition of the capacity accounting for memory effects, *i.e.,*

$$C^{\text{mem}} \triangleq \lim_{k\to\infty} \frac{1}{k} \sup_{X_k} I_{X_k,Y_k}\,, \tag{2.22}$$

where the MI between random vectors X_k and Y_k is defined as

$$I_{X_k,Y_k} \triangleq \mathbb{E}\left[\log \frac{p_{Y_k|X_k}(Y_k|X_k)}{p_{Y_k}(Y_k)}\right]. \tag{2.23}$$

However, if the input symbols are independent identically distributed (IID), the channel does not introduce any memory effects, and the *average* PDF is considered, then, Eq. (2.24) can be simplified to a memoryless channel capacity, which is equal to the maximum MI that can be achieved over all possible input distributions $p_X(x)$, *i.e.,*

$$C \triangleq \max_{p_X(x)} I_{X,Y} \leq C^{\text{mem}}\,, \tag{2.24}$$

where the multi-dimensional integral $I_{X,Y}$ denotes the MI function, and the maximisation must fulfil at least one constraint equation of the following form

$$\int_{\mathcal{X}} dx\, \rho(x)\, p_X(x) \leq P\,, \tag{2.25}$$

where $\rho(x)$ denotes a non-negative cost schedule function. Commonly, the input distribution $p_X(x)$ is subject to the average input power P constrained, which uses a quadratic cost schedule, *i.e.,* $\rho(x) = |x|^2$ so that

$$\mathbb{E}\left[|X|^2\right] \triangleq \int_{\mathcal{X}} dx\, |x|^2 p_X(x) \leq P\,. \tag{2.26}$$

Note that, in contract to MI, the channel capacity no longer depends on input distribution functions. In most cases, the capacity is a non-decreasing function of P, therefore, the inequality in Eq. (2.26) can be safely replaced by an equality. In spite of the MI functional is proven to be a concave with respect to $p_X(x)$, the optimisation over all possible distributions with respect to (2.26) remains a highly cumbersome problem. Nonetheless, for a Gaussian channel with inter-symbol interference (ISI) and average power constrained (2.26), the capacity (2.22) can be achieved if the input symbols X_k are correlated Gaussian RVs [36].

Further details about fundamental principles and metrics in information theory can be readily found in the pioneering work by Shannon [34], and in the following textbooks on information theory [32, 33, 37].

2.2 Signal propagation in optical fibre

2.2.1 Nonlinear optical susceptibility

Phenomenologically, an optical response of a dielectric material due to the interaction with electro-magnetic (EM) waves with a given strength $\mathbf{E}(\mathbf{r},t)$ can be effectively accounted by means of introducing an induced polarisation field $\mathbf{P}(\mathbf{r},t)$. Assuming weak NL behaviour of the medium, we can use the Tylor expansion of $\mathbf{P}(\mathbf{r},t)$ in power of the electric field $\mathbf{E}(\mathbf{r},t)$. This, in turn, allows us to split the induced field $\mathbf{P}(\mathbf{r},t)$ into its linear part $\mathbf{P}^{(\mathrm{lin})}(\mathbf{r},t)$ and NL part $\mathbf{P}^{(\mathrm{nl})}(\mathbf{r},t)$ as follows

$$\mathbf{P}(\mathbf{r},t) = \mathbf{P}^{(\mathrm{lin})}(\mathbf{r},t) + \mathbf{P}^{(\mathrm{nl})}(\mathbf{r},t), \tag{2.27}$$

where $\mathbf{r}$ and t stand for the position vector and time, respectively. Within the dipole approximation, the corresponding time-dependent components can be defined as follows[4]

$$P_i^{(\mathrm{lin})}(t) = \epsilon_0 \int_{-\infty}^{\infty} d\tau\, \chi_{ij}^{(1)}(t-\tau)\, E_j(\tau) = \epsilon_0 \chi_{ij}^{(1)}(t) * E_j(t), \tag{2.28}$$

and

$$\begin{aligned} P_i^{(\mathrm{nl})}(t) &= \epsilon_0 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\tau_1 d\tau_2\, \chi_{ijk}^{(2)}(t-\tau_1, t-\tau_2)\, E_j(\tau_1) E_k(\tau_2) \\ &\quad + \epsilon_0 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\tau_1 d\tau_2 d\tau_3\, \chi_{ijkl}^{(3)}(t-\tau_1, t-\tau_2, t-\tau_3)\, E_j(\tau_1) E_k(\tau_2) E_l(\tau_3) \\ &\quad + \ldots \triangleq P_i^{(2)}(t) + P_i^{(3)}(t) + \ldots, \end{aligned} \tag{2.29}$$

where the symbol "$*$" denotes a convolution product, ϵ_0 is the vacuum permittivity, and $P_i^{(2)}$ and $P_i^{(3)}$ are the contributions of the second- and third-order NL terms, respectively. All these convolution integrals reflect the fact of the finite time delay between optical response of the media and variation of the electric field $\mathbf{E}(\mathbf{r}, t)$. By applying the Fourier transform, these functions can be re-written in a frequency domain as

$$P_i^{(\mathrm{lin})}(\omega) = \chi_{ij}^{(1)}(\omega)\, E_j(\omega)\,, \tag{2.30}$$

and the second-order NL polarisation vector

$$\begin{aligned} P_i^{(2)}(\omega) &= \mathcal{F}\Big[P_i^{(2)}(t)\Big] \\ &= \frac{1}{2\pi} \int\int d\omega_1 d\omega_2\, \chi_{ijk}^{(2)}(\omega_1, \omega_1)\, E_j(\omega_1) E_k(\omega_2)\, \delta(\omega - \omega_1 - \omega_2), \end{aligned} \tag{2.31}$$

and third-order contribution can be expressed as

$$\begin{aligned} P_i^{(3)}(\omega) &= \mathcal{F}\Big[P_i^{(3)}(t)\Big] \\ &= \frac{1}{4\pi^2} \int\int\int d\omega_1 d\omega_2 d\omega_3\, \chi_{ijkl}^{(3)}(\omega_1, \omega_1)\, E_j(\omega_1) E_k(\omega_2) E_l(\omega_3) \\ &\quad \times \delta(\omega - \omega_1 - \omega_2 - \omega_3), \end{aligned} \tag{2.32}$$

[4]Henceforth we use the Einstein summation convention – the repeated indices are implicitly summed over the three Cartesian components, *i.e.,* $a_{ij}x_j \triangleq \sum_{j=1}^{3} a_{ij}x_j$.

where $\chi_{ij}^{(1)}$ is the linear optical susceptibility, which makes the prevalent contribution, $\chi_{ijk}^{(2)}$ and $\chi_{ijkl}^{(3)}$ are second- and the third-order NL optical susceptibility tensors, respectively. The last two tensors characterise the NL response of the dielectric media under the influence of applied electric field $\mathbf{E}$. Note that the impacts of both $\chi_{ijk}^{(2)}$ and $\chi_{ijkl}^{(3)}$ are much smaller in comparison to the linear $\chi_{ij}^{(1)}$ contribution.

Depending on the initial conditions, *i.e.,* the different pulse shapes injected into the optical fibre, the $\chi^{(3)}$ nonlinearity gives rise to various NL optical effects. Specifically, $\chi^{(3)}$ term is responsible for the NL effects, such as third-order harmonic generation and the *optical* Kerr effect. The last will be explained in detail in Sec. 2.2.2. It should be emphasised that in silica-glass fibre, the second-order term can be neglected due to its inversion symmetry at the molecular level of SiO_2. Excepting for the relatively weak electric-quadrupole and magnetic-quadrupole moments, there is no second-order NL effects arising from $\chi^{(2)}$ in optical fibre. It can be claimed that $\chi^{(3)}$ is the main source of nonlinearity in SSMF without any loss of generality. Generally speaking, the 4^{th} -rank tensor $\chi_{ijkl}^{(3)}$ has $3^4 = 81$ complex components. The real parts $\mathrm{Re}\left[\chi^{(3)}\right]$ are responsible for the Kerr effects, whereas its imaginary parts $\mathrm{Im}\left[\chi^{(3)}\right]$ are mainly for the NL Raman scattering process. The Raman scattering effect, however, becomes much more significant for shorter pulses, *i.e.,* $\Delta\tau \lesssim 10\,\mathrm{ps}$. Thus, we can stress that the most relevant effects appearing in optical fibre: an NL refraction (*i.e.,* self-focusing) as well as an NL mixing (*i.e.,* four-wave mixing (FWM)), both are consequently described by the real components of $\chi^{(3)}$.

2.2.2 Optical Kerr effect

The influence of $\chi^{(3)}$ makes the *effective* refractive index dependent on the optical intensity I

$$n(\omega, I) = n_0(\omega) + n_2 I\,, \tag{2.33}$$

where n_2 is the optical Kerr coefficient. Taking into account that

$$I = \frac{\langle |\mathbf{E}|^2 \rangle_t}{Z}\,, \tag{2.34}$$

where $Z = \sqrt{\mu_0 / (\epsilon_0 \epsilon)}$ is the so-called wave impedance, μ_0 is the vacuum permeability, $\langle \cdot \rangle_t$ stands for a time average operator, the equation (2.33) can be

readily re-written in terms of electric field as

$$n(\omega, |\mathbf{E}|^2) = n_0(\omega) + \tilde{n}_2 |\mathbf{E}|^2 , \tag{2.35}$$

where the spectrum of linear refractive index $n_0(\omega)$ is straightforwardly referring to the spectrum of linear optical susceptibility $\chi^{(1)}(\omega)$

$$n_0^2(\omega) = 1 + \chi^{(1)}(\omega) , \tag{2.36}$$

and it also can be expressed via the well-known Sellmeier approximation [38, 39]

$$n_0^2(\omega) - 1 = \sum_{l=1}^{k} \frac{B_l \omega^2}{\omega^2 - \omega_l^2} , \tag{2.37}$$

where the summation is taken over all resonances in a fibre, which contribute to the given frequency range, and the coefficients B_l refer to the strength of l-absorption resonance at the frequency ω_l, and this is usually determined from experimental data, and called the Sellmeier coefficients. For SSMF, the Sellmeier coefficients can be obtained by fitting experimentally measured dispersion curve with setting $k = 3$ in Eq. (2.37). For bulk-fused silica at 20 °C, at wavelengths $\lambda_1 = 68.4043\,\text{nm}$, $\lambda_2 = 116.2414\,\text{nm}$, and $\lambda_3 = 9.896\,161\,\mu\text{m}$, the corresponding parameters are found to be equal to $B_1 = 0.6961663$, $B_2 = 0.4079426$, and $B_3 = 0.8974794$ [40].

The relationship between NL refractive index n_2 in (2.33) and $\tilde{n}_2$ in (2.35) is given by

$$\frac{2\tilde{n}_2}{n_2} = c\epsilon_0 n_0 . \tag{2.38}$$

It should be noted that the NL refractive index $\tilde{n}_2$ has a direct connection with the real part of the aforementioned third-order NL optical susceptibility $\chi^{(3)}$. Assuming a linearly polarised optical electric field $\mathbf{E}$, *i.e.,* neglecting the NL birefringence due to the tensorial nature of the fourth rank tensor $\chi^{(3)}_{ijkl}$, therefore, only one component $\chi^{(3)}_{1111}$ makes the contribution, *i.e.,*

$$\tilde{n}_2 = \frac{3}{8n_0} \operatorname{Re}\left[\chi^{(3)}_{1111}\right] . \tag{2.39}$$

For SSMF, at $\lambda_0 = 1550$ nm, the linear refraction index $n_0 = 1.46$, and the real part of the third-order optical susceptibility is $1.56 \times 10^{-19}\,\text{W}^{-1}$, therefore, the value of NL refractive index n_2 in Eq. (2.33) is typically of the order of $10^{-20}\,\text{m}^2\,\text{V}^{-2}$ (see, *e.g.,* [17, 41]). Within the telecommunication window, an approximate value $2.35 \times 10^{-20}\,\text{m}^2\,\text{V}^{-2}$ was measured for pure silica (without

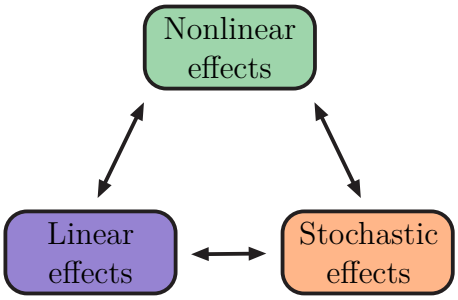

Figure 2.1: Physical phenomena in optical fibre classified in three groups: nonlinear effects: SFM, XPM, FWM, self-steepening, Raman scattering); linear effects (chromatic dispersion, dispersion slope); Stochastic effects (ASE noise, short noise, thermal noise).

doping) fibres [42].

2.2.3 Generalised nonlinear Schrödinger equation

The aim of this section is to outline the mathematical description for the optical fibre waveforms channel. The propagation of electric field $\mathbf{E}$ in any dielectric media obeys a NL wave equation, which can be straightforwardly derived from the system of Maxwell equations assuming the absence of free charges (*i.e.*, $\nabla \cdot \mathbf{D} = \epsilon\epsilon_0 \nabla \cdot \mathbf{E} = 0$) in a non-conducting (*i.e.*, $\mathbf{J} = 0$) and non-magnetic (*i.e.*, $\mathbf{M} = 0$, so that $\mathbf{B} = \mu_0 \mathbf{H}$) medium (for details, see [43]). We then have

$$\Box \mathbf{E}(\mathbf{r}, t) = -\mu_0 \frac{\partial^2 \mathbf{P}(\mathbf{r}, t)}{\partial t^2}, \tag{2.40}$$

where $\Box \triangleq \left(\partial^2 / c^2 \partial t^2\right) - \triangle$ stands for the d'Alembertian, and $\triangle$ is the Laplacian, and c is the speed of light in vacuum. Separating the linear and NL parts the induced polarisation vector (polarisability) $\mathbf{P}$ according to (2.27), and then removing all second-order NL contributions, the dipole moment of $\mathbf{P}$ yields

$$\mathbf{P}(\mathbf{r}, t) \approx \epsilon_0 \left(\chi^{(1)} \mathbf{E}(\mathbf{r}, t) + \chi^{(3)} |\mathbf{E}(\mathbf{r}, t)|^2 \mathbf{E}(\mathbf{r}, t)\right), \tag{2.41}$$

where $\chi^{(3)}$ is responsible for fibre NL effects (see Eqs (2.35), (2.39)). The key assumption in derivation of NLSE is that the optical signal bandwidth is supposed to be small in comparison with the carrier frequency ω_0 which, in turn, gives a permit to make a decomposition into a slowly-varying amplitude and a highly-oscillating carrier term $e^{\imath\omega_0 t}$, which is widely known as the slowly varying envelope approximation (SVEA). For the carrier wavelength $\lambda_0 = 1550\,\text{nm}$, the SVEA is fulfilled for an optical pulse duration $\gtrsim 1\,\text{ps}$.

Since the fibre core size is assumed to be greater than the optical wave-

length, the longitudinal component of the incident electric field is sufficiently small and can be ignored compared to the transverse components. Thus, we seek for the solution of Eq. (2.40) in the trial form of a *guided wave*, *i.e.,*

$$\mathbf{E}(z,t) = \hat{\mathbf{n}}\, F(x,y)\, Q(z,t) \exp\left[\imath\left(\beta_0 z - \omega_0 t\right)\right], \tag{2.42}$$

where $Q(z,t)$ is the slowly varying envelope of electric field $E(z,t)$, the function $F(x,y)$ represents the transverse distribution inside the fibre core, $\beta_0 = \omega_0/c$ is a mode-propagation (phase) constant, and $\hat{\mathbf{n}}$ is the unit vector (*i.e.,* $|\hat{\mathbf{n}}|$=1), which is assumed to be parallel to the vector $\mathbf{P}(\mathbf{r},t)$.

Combining Eqs. (2.35), (2.40), (2.42), and assuming that the intensity dependence of $n(\omega)$ is frequency independent, we can obtain the following partial differential equation of a slowly varying envelope signal, which is referred to as a scalar NLSE [44, 45][5]

$$\underbrace{\imath\frac{\partial Q(z,t)}{\partial z}}_{\text{Spatial signal evolution}} \underbrace{+\imath\frac{\alpha}{2}Q(z,t) - \frac{\beta_2}{2}\frac{\partial^2 Q(z,t)}{\partial t^2} + \imath\frac{\beta_3}{6}\frac{\partial^3 Q(z,t)}{\partial t^3}}_{\text{Linear effects: fibre loss and dispersion}} + \underbrace{\gamma\,|Q(z,t)|^2 Q(z,t)}_{\text{Kerr nonlinearity}} = 0, \tag{2.43}$$

where z denotes the propagation distance measured in [km], assuming that the transmitter is located at $z=0$ and the receiver is at $z=L$, where L is the total transmission distance; t is the retarded time measured in [s], and defined as $t = \tilde{t} - \beta_1 z$, where α is the fibre loss parameter, β_2 and β_3 represent the second- and third-order dispersion coefficients, respectively, and γ is the nonlinearity parameter (a measure of Kerr nonlinearity in the fibre):

$$\gamma = \frac{2\pi}{\lambda_0}\frac{n_2}{A_{\text{eff}}}, \tag{2.44}$$

where the *effective mode area* of the fibre is defined as

$$A_{\text{eff}} = \frac{\left(\int_0^\infty \int_0^\infty |F(x,y)|^2\, dxdy\right)^2}{\int_0^\infty \int_0^\infty |F(x,y)|^4\, dxdy}, \tag{2.45}$$

which essentially represents the area of the fibre core would have if the optical power is assumed to be uniformly distributed across the fibre core, for instance, typically, $A_{\text{eff}} \sim 80\,\mu\text{m}^2$ for SSMF. Eq. (2.43) was derived, for the first time,

[5]We do not consider the polarisation evolution of the $\mathbf{E}(\mathbf{r},t)$, *i.e.,* the polarisation mode dispersion (PMD) effect is not considered.

by Hasegawa and Tappert [46, 47]. The detailed derivations are also provided in [45].

In the presence of additive ASE noise Eq. (2.43) can be generalised to the so-called *stochastic* scalar NLSE[6] [48]

$$\imath\frac{\partial Q(z,t)}{\partial z}+\imath\frac{\alpha}{2}Q(z,t)-\frac{\beta_2}{2}\frac{\partial^2 Q(z,t)}{\partial t^2}+\gamma\left|Q(z,t)\right|^2 Q(z,t)=N(z,t), \tag{2.46}$$

where $N(z,t)$ is the circularly symmetric white (both in space and in time) Gaussian noise with the following correlation properties

$$\begin{aligned}&\mathbb{E}\left[N(z,t)\right]=\mathbb{E}\left[N(z,t)\,N(z',t')\right]=0,\\&\mathbb{E}\left[N(z,t)\,N^*(z',t')\right]=\sigma_0^2\,\delta\left(z-z'\right)\delta\left(t-t'\right),\end{aligned} \tag{2.47}$$

where σ_0^2 is the power spectral density (PSD) of ASE noise per km, $\mathbb{E}\left[\cdot\right]$ is the mathematical expectation operator, and $\delta\left(\cdot\right)$ is the Dirac delta function.

It is worth mentioning that the propagation of an optical pulse with a duration less than $\lesssim 1\,\mathrm{ps}$ gives rise to the additional terms, which are responsible for the effects of self-steepening and Raman scattering, and the chromatic dispersion slope, *i.e.,*

$$\begin{aligned}&\imath\frac{\partial Q(z,t)}{\partial z}+\imath\frac{\alpha}{2}Q(z,t)-\frac{\beta_2}{2}\frac{\partial^2 Q(z,t)}{\partial t^2}+\imath\frac{\beta_3}{6}\frac{\partial^3 Q(z,t)}{\partial t^3}+\gamma\left|Q(z,t)\right|^2 Q(z,t)\\&+\imath\gamma s_0\frac{\partial\left(\left|Q(z,t)\right|^2 Q(z,t)\right)}{\partial t}-\gamma T_{\mathrm{R}}\,Q(z,t)\frac{\partial\left|Q(z,t)\right|^2}{\partial t}=N(z,t),\end{aligned} \tag{2.48}$$

where the term captured by $s_0=\lambda_0/(\pi c)$ is responsible for the self-steepening effect of the pulse edge[7], where λ_0 denotes the optical centre wavelength. The term with a pre-factor T_{R} is related to the Raman scattering effect, where the characteristic time T_{R} is connected to to the slope of the Raman gain spectrum. The solution of Eq. (2.48) inherently represent the interaction among the stochastic effects (*i.e.,* noisy term on the right-hand side $N(z,t)$), the linear effects (chromatic dispersion captured by β_2, the dispersion slope captured by β_3, and the fibre attenuation α), and the fibre nonlinearity (captured by the term with a pre-factor γ) due to the above-mentioned optical Kerr effect. All these effects and their interactions are schematically visualised in Fig. 2.1. These effects can be distinct as well as can influence each other.

[6]Here the third-order dispersion effect (*i.e.,* a dispersion slope) as well as higher order dispersion contributions are omitted.

[7]In optical communications the effect of self-steepening can be safely neglected.

2.2.4 Manakov equation

In the previous section, the stochastic NLSE Eq. (2.46) has been introduced as a master model for single polarisation propagation. For dual polarisation (DP) case, the propagation of slowly varying transverse components $Q_x(z,t)$ and $Q_y(z,t)$ through a fibre obeys the coupled-mode NLSE [49, 50]

$$\imath\frac{\partial Q_x(z,t)}{\partial z}+\imath\frac{\alpha}{2}Q_x(z,t)-\frac{\beta_2}{2}\frac{\partial^2 Q_x(z,t)}{\partial t^2}+\gamma\left(|Q_x(z,t)|^2+\frac{2}{3}|Q_y(z,t)|^2\right)Q_x(z,t)+\frac{\gamma}{3}Q_y^2(z,t)\,Q_x^*(z,t)=N_x(z,t)\,, \tag{2.49}$$

and

$$\imath\frac{\partial Q_y(z,t)}{\partial z}+\imath\frac{\alpha}{2}Q_y(z,t)-\frac{\beta_2}{2}\frac{\partial^2 Q_y(z,t)}{\partial t^2}+\gamma\left(|Q_y(z,t)|^2+\frac{2}{3}|Q_x(z,t)|^2\right)Q_x(z,t)+\frac{\gamma}{3}Q_x^2(z,t)\,Q_y^*(z,t)=N_y(z,t)\,, \tag{2.50}$$

Eqs. (2.49) and (2.50) can be re-written in a form of vector equation as follows [51, Eq. (2)]

$$\begin{aligned}&\imath\frac{\partial\mathbf{Q}(z,t)}{\partial z}+\imath\frac{\alpha}{2}\mathbf{Q}(z,t)-\frac{\beta_2}{2}\frac{\partial^2\mathbf{Q}(z,t)}{\partial t^2}\\&+\gamma\left[\mathbf{Q}(z,t)\,\mathbf{Q}^\dagger(z,t)\,\mathbf{Q}(z,t)-\frac{1}{3}\left(\mathbf{Q}^\dagger(z,t)\,\hat{\boldsymbol{\sigma}}_y\,\mathbf{Q}(z,t)\right)\hat{\boldsymbol{\sigma}}_y\,\mathbf{Q}(z,t)\right]\\&=\mathbf{N}(z,t)\,,\end{aligned} \tag{2.51}$$

where $\mathbf{Q}(z,t)=\begin{bmatrix}Q_x(z,t)\\Q_y(z,t)\end{bmatrix}$ is the Jones vector, $\hat{\boldsymbol{\sigma}}_y=\begin{pmatrix}0&-\imath\\\imath&0\end{pmatrix}$ is one of the Pauli spin matrices, and $\mathbf{N}(z,t)=[N_x(z,t),N_y(z,t)]^{\mathrm{T}}$, where $N_x(z,t)$ and $N_y(z,t)$ are independent and identically distributed noise terms, "†" denotes the Hermitian transpose. Eq. (2.51) is valid in the absence of any birefringence in a fibre. However, in real optical fibre, the birefringence changes randomly and cannot be neglected. By averaging over rapidly changing random polarisation states, one can obtain the so-called Manakov equation [52, 53]

$$\imath\frac{\partial\mathbf{Q}(z,t)}{\partial z}+\imath\frac{\alpha}{2}\mathbf{Q}(z,t)-\frac{\beta_2}{2}\frac{\partial^2\mathbf{Q}(z,t)}{\partial t^2}+\frac{8\gamma}{9}\mathbf{Q}(z,t)\,\mathbf{Q}^\dagger(z,t)\,\mathbf{Q}(z,t)=\mathbf{N}(z,t)\,. \tag{2.52}$$

It should be noted that, in contrast to coupled NLSEs, both polarisations of $\mathbf{Q}(z,t)$ in Eq. (2.52) have the same NL phase shift.

2.2.5 Fibre attenuation effect

Accounting for the effect of fibre loss effect only leads a straightforward integration by using a method of separation of variables. In this particular case Eq. (2.43) has the following simple form

$$\frac{\partial Q(z,t)}{\partial t} + \frac{\alpha}{2} Q(z,t) = 0\,, \tag{2.53}$$

and after the propagation over distance z, the solution can be readily obtained

$$Q(z,t) = Q(0,t) \exp\left(-\frac{\alpha z}{2}\right), \tag{2.54}$$

where the envelope of the electric field $Q(z,t)$ decays exponentially with distance, and α represents the attenuation (fibre loss) parameter. Practically, it is convenient to rewrite Eq. (2.53) in terms of optical power $P(z,t)$ (also known as the *Beer's law*)

$$\frac{\partial P(z,t)}{\partial t} = -\alpha P(z,t)\,, \tag{2.55}$$

where $P(z,t) \triangleq |Q(z,t)|^2$ is the instantaneous optical power. The solution of (2.53) is also obvious, so that the fibre loss effect can be easily modelled as follows

$$P(z) = P_0\, e^{-\alpha z}\,, \tag{2.56}$$

where $P_z \triangleq \langle P(z,t)\rangle_t$ is the power profile, $P_0 \triangleq P(z,t)|_{z=0} = P(0,t)$ denotes the launched (input) power of the optical signal. It is often convenient to express units of parameter α in dB per km, *i.e.,*

$$\alpha\left[\frac{\text{dB}}{\text{km}}\right] = -\frac{10}{L} \lg \frac{P(L)}{P_0} = (10 \lg e) \cdot \alpha \approx 4.343\, \alpha\,. \tag{2.57}$$

For SSMF and the wavelengths around $\lambda_0 = 1550$ nm, $\alpha \approx 0.20$ dB/km (see Fig. 2.2).

Physically speaking, we can distinguish two main attenuation mechanisms in fibre, such as the *infra-red absorption* process and the *Rayleigh scattering*. These are precisely two effects that determine the minimum loss, which, in turn, corresponds to the optical wavelength $\lambda_0 = 1550$ nm. Formally, the presence of loss implies that a longitudinal component of propagation constant is no longer a real function, and has an imaginary part. The absorption may come from both absorption of the material (*i.e.,* in silica) as well as the im-

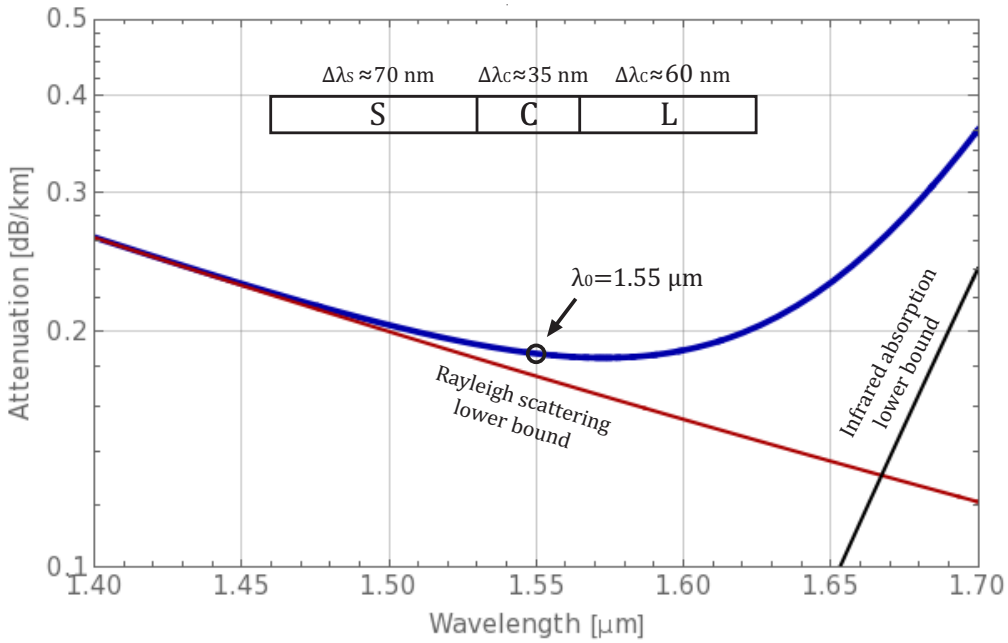

Figure 2.2: Fibre attenuation α as a function of wavelength based on the model in (2.58).

purities in optical fibre. Material absorption can include electronic resonances in the ultraviolet region and molecular resonances in the far-infra-red region. Note that, ultraviolet absorption is assumed to be sufficiently small in the low-loss window. There is also an absorption peak around the wavelength 1390 nm which is due to the impurities (OH^{-1} ions) in glasses.

Ignoring all losses, which are due to the imperfections in fibre, the total attenuation coefficient can be approximately modelled as [54]

$$\alpha \approx \alpha_{\mathrm{R}} + \alpha_{\mathrm{IR}}\,, \tag{2.58}$$

where α_{R} and α_{IR} represent the contributions of the Rayleigh scattering and infra-red absorption, respectively.

In silica-based fibres, the attenuation within the wavelength range from 1300 nm to 1600 nm is essentially lower-bounded by Rayleigh scattering process, which emerges from microscopic fluctuations in the optical density of the medium, and thus, sets the fundamental limit. Note that, glasses (or meta-materials) with attenuation lower than the Rayleigh scattering limits are still unknown. The Rayleigh scattering term α_{R} of fibre attenuation parameter α decreases rapidly with increasing optical wavelength λ, and the approximate

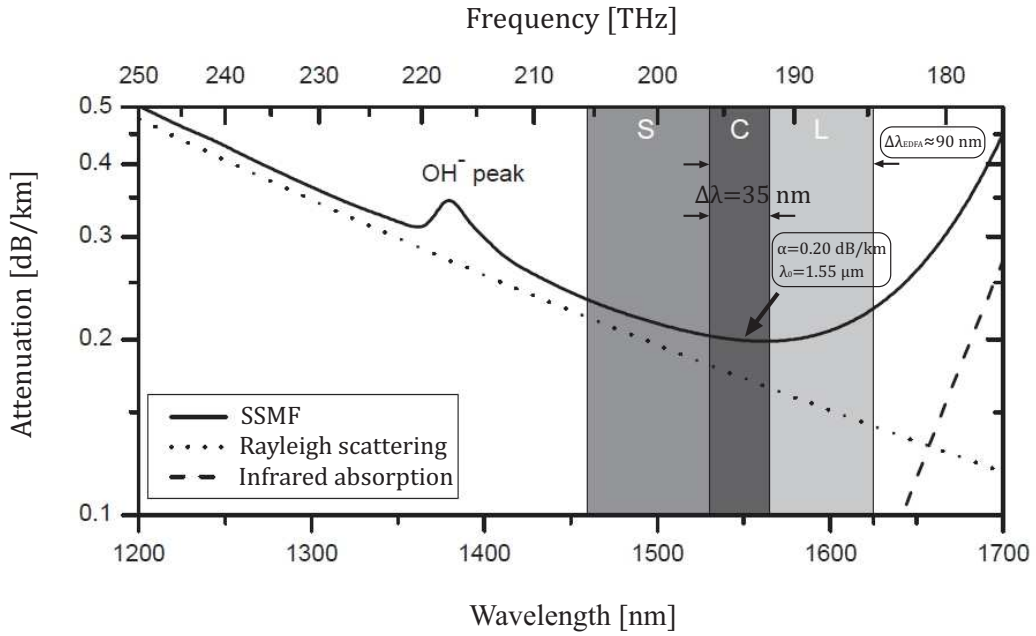

Figure 2.3: A typical SSMF attenuation profile as a function of both wavelength and frequency.

wavelength dependence can be given by

$$\alpha_{\mathrm{R}} = \frac{A}{\lambda^4}\,, \tag{2.59}$$

where A stands for the Rayleigh scattering coefficient. The most commonly used model for the Rayleigh coefficient is given by the following closed-form expression [55]

$$A = \frac{8\pi^3}{3}\, n_0^8 p^2 \beta_{\mathrm{T}} \cdot k_{\mathrm{B}} T_{\mathrm{F}}\,, \tag{2.60}$$

where n_0 is the linear refractive index, p is the average photo-elastic coefficient, β_{T} is the isothermal compressibility, T_{F} is the fictive temperature, k_{B} is the Boltzmann's constant. The fictive temperature represents the temperature at which no structural relaxation occurs in glasses and it changes depending mostly on the glass viscosity. Typical values of these coefficients for pure silica are given in Tab. 2.1.

For pure silica glass, the Rayleigh scattering coefficient is approximately given by $A \approx 0.8\,\mathrm{dB\,\mu m^4\,km^{-1}}$, whereas for a SSMF this value is always slightly greater $A \approx 1.02\,\mathrm{dB\,\mu m^4\,km^{-1}}$.

For the wavelengths above 1600 nm, the infra-red absorption becomes

Table 2.1: Typical property of silica fibre

Parameters	Values
n_0	1.49
p	0.286
β_{T}	$7.25 \times 10^{11}\,\mathrm{Pa}^{-1}$
T_{F}	1433 K

Table 2.2: Usable wavelength bands in optical fibre

Band	Description	Approximate wavelength range
S-band	Short band	1460 nm to 1530 nm
C-band	Conventional band	1530 nm to 1565 nm
L-band	Long band	1565 nm to 1625 nm

relevant. It can be well-described by following model

$$\alpha_{\mathrm{IR}} = C \exp\left(-\frac{D}{\lambda}\right), \tag{2.61}$$

where C and D are absorption coefficients which refer to property of the material. Typically, for pure silica, $C \approx 6.65 \times 10^{12}\,\mathrm{dB\,km^{-1}}$ and $D \approx 52.62\,\mu\mathrm{m}$.

It should also be emphasised that the usable bandwidth of optical fibre is substantially restricted by the bandwidth of available optical amplifiers rather than the bandwidth of silica-based fibre itself. For instance, the bandwidth availability of modern optical amplifiers (*e.g.,* EDFA, optical Raman amplifier, Hybrid), one can distinguish three optical bands such as S-band, C-band and L-band (see Tab. 2.2) within the low-loss region in fibre around a reference centre wavelength $\lambda_0 = 1.55\,\mu\mathrm{m}$.

The C-band is attributed to the conventional EDFAs and provides a channel bandwidth approximately 5 THz, whereas the L-band is currently being used for high-capacity WDM transmission systems by means of shifting the gain of EDFAs to provide an optical signal amplification within this band. As shown in Fig. 2.3, the total EDFA bandwidth is roughly equal to $\Delta\lambda_{\mathrm{EDFA}} \approx 90\,\mathrm{nm}$.The amplification within S-band can be provided by using Raman amplifiers. Some novel fibres have been recently fabricated to open up the opportunity of using a new low-loss window called E-band (extended band) with a wavelength range from 1360 nm to 1460 nm. Nevertheless, there are still no appropriate amplifiers have been created to supply an optical amplification within this band.

2.2.6 Fibre dispersion effect

Neglecting optical Kerr effects in fibre, the equation can be written as follows

$$\imath \frac{\partial Q(z,t)}{\partial z} - \frac{\beta_2}{2} \frac{\partial^2 Q(z,t)}{\partial t^2} = 0\,. \tag{2.62}$$

Evidently, this equation can be solved straightforwardly by using the conventional Fourier transform. The solution can be analytically represented as a convolution of envelope of input electric field with an impulse response function of a "dispersion filter". Applying Fourier transform, Eq. (2.62) yields

$$\frac{\partial U(z,\omega)}{\partial z} = \imath \frac{\beta_2 \omega^2}{2} U(z,\omega)\,, \tag{2.63}$$

where $\omega = 2\pi\nu$ is the angular frequency, and $U(z,\omega) \triangleq \mathcal{F}\left[Q(z,t)\right]$ stands for the Fourier amplitude. Separating the variable in Eq. (2.63), it yields

$$U(z,\omega) = H(z,\omega)\, U(0,\omega)\,, \tag{2.64}$$

where $H(z,\omega)$ is the $z-$dependent frequency response of the dispersion filter, which is defined as

$$H(z,\omega) = \exp\left(\imath \frac{\beta_2 \omega^2}{2} \cdot z\right). \tag{2.65}$$

Applying the inverse Fourier transform to Eq. (2.64), the solution of Eq. (2.62) in the time domain is

$$Q(z,t) = h(z,t) * Q(0,t)\,, \tag{2.66}$$

where $''*''$ denotes the convolution, and $h(z,t) \triangleq \mathcal{F}^{-1}\left\{H(z,\omega)\right\}$ is the impulse response of the dispersive filter (all-pass filter), which is given by

$$h(z,t) = \frac{1}{\sqrt{2\pi\imath\beta_2 z}} \exp\left(-\frac{t^2}{2\imath\beta_2 z}\right). \tag{2.67}$$

It is easy to see that Eq. (2.67) is essentially a Gaussian function. Therefore, the effect of chromatic dispersion in fibre can effectively be modelled as an all-pass Gaussian filter with the impulse response Eq. (2.67). It should also be noted that the effect of chromatic dispersion does not change the signal spectrum amplitude, however, it gives rise to a frequency-dependent phase shift $\beta_2(\omega)\,\omega^2/2$.

Notably, the effect of chromatic dispersion leads to a frequency-dependence of a group velocity v_g (*i.e.,* a speed of pulse envelope). In other words, different frequency components of the optical pulse move with different

group velocities, which is also known as a group velocity dispersion (GVD). After propagation over distance L, the signal frequency components consequently have a delay L/v_g that, in turn, leads a pulse broadening in the time domain. The pulse broadening ΔT can be readily expressed as

$$\Delta T = \frac{d}{d\omega}\frac{L}{v_g}\Delta\omega = \beta_2 L\,\Delta\omega\,, \tag{2.68}$$

where $\Delta\omega$ is the spectral width of the pulse.

Generally speaking, the total electric field envelope after propagation along distance z experiences the a frequency-dependent phase shift $\beta(\omega)z$, and is given by

$$Q(z,t) = \int_{-\infty}^{\infty} d\omega\, U(0,\omega)\exp\Big[-\imath\,(\omega t - \beta(\omega)z)\Big]\,, \tag{2.69}$$

where the propagation constant $\beta(\omega)$ is connected with the refractive index $n(\omega)$ as follows

$$\beta(\omega) = \frac{\omega}{c}\,n(\omega)\,, \tag{2.70}$$

and can be expanded in a Taylor series about the carrier frequency ω_0 as

$$\beta(\omega) = \beta_0 + \beta_1\,(\omega-\omega_0) + \frac{\beta_2}{2}\,(\omega-\omega_0)^2 + \frac{\beta_3}{3!}\,(\omega-\omega_0)^3 + \ldots\,, \tag{2.71}$$

with $\beta_0 \triangleq \beta(\omega_0)$, and with

$$\beta_1 \triangleq \frac{d\beta}{d\omega} = \frac{1}{c}\,\frac{d}{d\omega}\,\omega n(\omega) = \frac{1}{c}\left(n + \omega\frac{dn}{d\omega}\right). \tag{2.72}$$

Hence, using Eq. (2.72), the group velocity v_g in Eq. (2.68) can be readily expressed as follows [56]

$$v_g \triangleq \frac{1}{\beta_1} = \frac{c}{n_g}\,, \tag{2.73}$$

where n_g represents the so-called group refractive index given by

$$n_g \triangleq n + \omega\frac{dn}{d\omega}\,. \tag{2.74}$$

The coefficient β_2 characterises the GVD (*i.e.,* second-order dispersion), *i.e.,*

$$\beta_2 \triangleq \frac{d^2\beta}{d\omega^2} = \frac{1}{c}\,\frac{dn_g}{d\omega} = -\frac{1}{v_g^2}\,\frac{dv_g}{d\omega} = \frac{1}{c}\left(2\frac{dn}{d\omega} + \omega\frac{d^2n}{d\omega^2}\right). \tag{2.75}$$

It is customary, especially in the theory of optical communication systems, to use a wavelength domain, *i.e.,*

$$\Delta\omega = -\frac{2\pi c}{\lambda_0^2}\Delta\lambda\,. \tag{2.76}$$

Hence, a group velocity delay ΔT in (2.68) between two channels with a channel spacing in wavelength domain $\Delta\lambda$ can be equivalently expressed as follows

$$\Delta T = \frac{d}{d\lambda}\frac{L}{v_g}\Delta\lambda = DL\,\Delta\lambda\,, \tag{2.77}$$

where D is the chromatic dispersion parameter, which is at the centre wavelength λ_0 related to β_2 as follows

$$D \triangleq \left(\frac{d\beta_1}{d\lambda}\right)_{\lambda=\lambda_0} = \frac{1}{c}\frac{dn_g}{d\lambda} = -\frac{\lambda_0}{c}\frac{d^2n}{d\lambda^2} = -\frac{2\pi c}{\lambda_0^2}\beta_2\,, \tag{2.78}$$

where the group refractive index n_g is now expressed in a wavelength domain as follows

$$n_g = n - \lambda\frac{dn}{d\lambda}\,. \tag{2.79}$$

Using (2.68) and (2.78), we can obtain a very simple and useful relationship between the second-order dispersion coefficient β_2 and the group delay parameter D, *i.e.,*

$$\beta_2\Delta\omega = D\Delta\lambda\,. \tag{2.80}$$

Higher order dispersion effects give rise to a dispersion slope S, *i.e.,*

$$S \triangleq \left(\frac{dD}{d\lambda}\right)_{\lambda=\lambda_0} = \frac{(2\pi c)^2}{\lambda_0^4}\beta_3 + \frac{4\pi c}{\lambda_0^3}\beta_2\,, \tag{2.81}$$

where the second-order β_2 and the third-order β_3 dispersion coefficients in Eq. (2.71) can be expressed in terms of wavelength through

$$\beta_3 = \frac{\lambda^3}{2\pi c^2}\frac{d^2n}{d\lambda^2}\,, \tag{2.82}$$

$$\beta_3 \triangleq \frac{d^3\beta}{d\omega^3} = -\frac{\lambda^4}{4\pi^2c^3}\left(3\frac{d^2n}{d\lambda^2} + \lambda\frac{d^3n}{d\lambda^3}\right). \tag{2.83}$$

Obviously, the dispersion slope has higher impact for transmission in the region of the zero-dispersion wavelength λ_{ZDW} at which $D = 0$. Practically, such an effect occurs when the material dispersion and wave-guide disper-

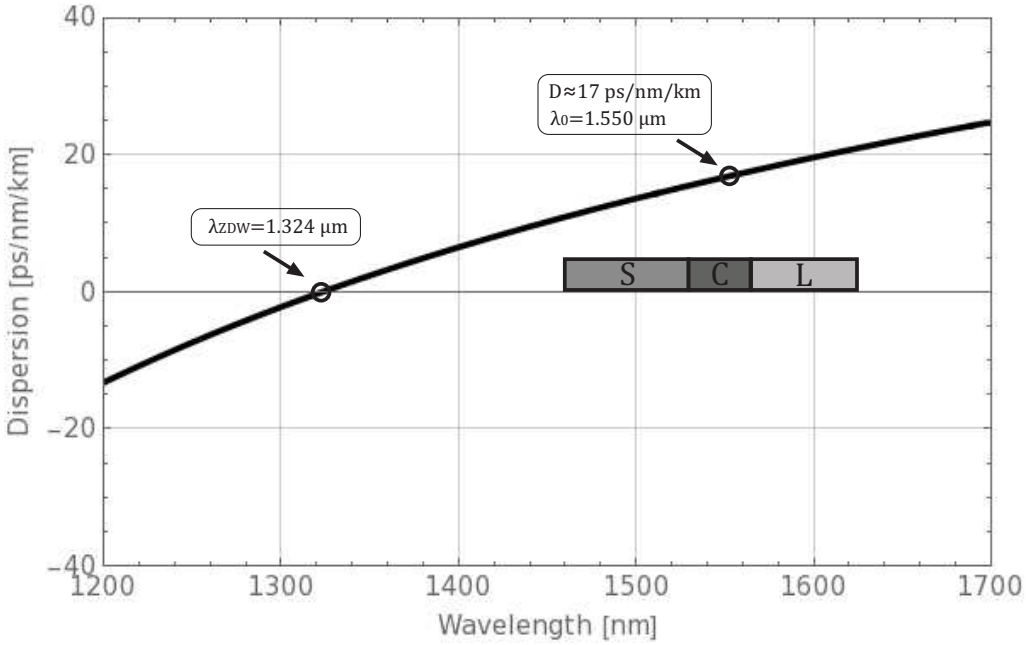

Figure 2.4: Chromatic dispersion parameter D for typical SSMF versus wavelength based on the approximation Eq. (2.84).

sion cancel each other out. In pure silica, $\lambda_{\mathrm{ZDW}} \approx 1310\,\mathrm{nm}$. Below the zero-dispersion wavelength λ_{ZDW}, pulses in optical fibre experience *normal* chromatic dispersion (*i.e.*, $\beta_2 > 0$ or $D < 0$) – longer wavelengths travel faster. However, in the entire 1550 nm band in SSMF, shorter wavelengths optical pulses travel faster experiencing *anomalous* dispersion (*i.e.*, $\beta_2 < 0$ or $D > 0$).

In SSMF for the operating wavelength range 1200 nm to 1600 nm, the wavelength dependence of the dispersion parameter D measured in $\left[\mathrm{ps\,nm^{-1}\,km^{-1}}\right]$ can be approximated by the following empirical relationship

$$D(\lambda) \approx \frac{S_0}{4}\left(\lambda - \frac{\lambda_{\mathrm{ZDW}}^4}{\lambda^3}\right), \tag{2.84}$$

where $S_0 \leq 0.092\,\mathrm{ps\,nm^{-2}\,km^{-1}}$ represents the zero dispersion slope, and $1302\,\mathrm{nm} \leq \lambda_{\mathrm{ZDW}} \leq 1324\,\mathrm{nm}$ is the zero dispersion wavelength range. Approximated Eq. (2.84) is plotted in Fig. 2.4.

It should be noted that the overall optical fibre dispersion is determined by the sum of the material D_M and the waveguide dispersion D_W (Fig. 2.5), *i.e.*,

$$D = D_M + D_W\,, \tag{2.85}$$

where the material dispersion D_M is essentially related to the derivative of the

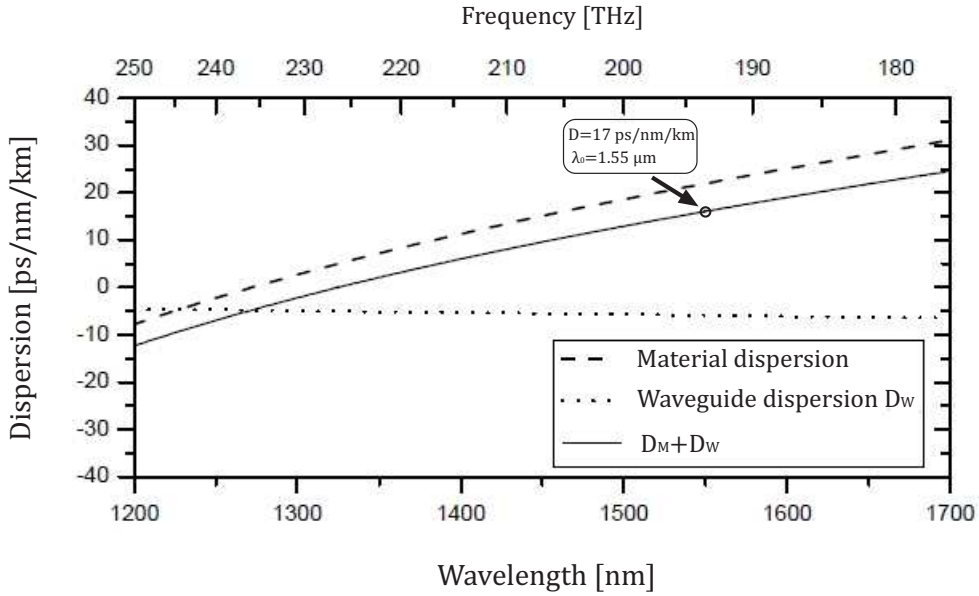

Figure 2.5: Chromatic dispersion parameter D for typical SSMF as a function of wavelength and frequency.The relative contributions of material dispersion D_M and waveguide dispersion D_W are also plotted.

group the group refractive index in the wavelength domain, *i.e.,*

$$D_M = \frac{1}{c}\frac{dn_g}{d\lambda}, \tag{2.86}$$

whereas the waveguide dispersion D_W depends on the core radius and the index difference between the core and the cladding. In other words, the waveguide dispersion can be modified by changing the refractive index profile. Fig. 2.5 indicates that D_M monotonically decreases with λ and equals 0 at $\lambda \approx 1276$ nm, whilst D_W decreases with λ and is always negative.

2.2.7 Fibre nonlinearity effect

Another special case of NLSE is the case of the absence of chromatic dispersion effect. In this particular case Eq. (2.46) reads

$$\imath\frac{\partial Q(z,t)}{\partial z} + \imath\frac{\alpha}{2}Q(z,t) + \gamma\,|Q(z,t)|^2\,Q(z,t) = 0\,, \tag{2.87}$$

by setting the transformation

$$Q(z,t) \doteq \exp\left(-\alpha z/2\right)\,Q(z,t), \tag{2.88}$$

we have

$$\imath\frac{\partial Q(z,t)}{\partial z} + \gamma\,|Q(z,t)|^2\,Q(z,t)\,e^{-\alpha z} = 0\,. \tag{2.89}$$

After the propagation distance L, the solution can be expressed as

$$Q(z,t) = Q(0,t)\exp\left[\imath\gamma\left|Q(0,t)\right|^2 \int_0^L e^{-\alpha z'}\, dz'\right]. \tag{2.90}$$

Denoting the launched peak power as $P_0 \triangleq |Q(0,t)|^2$ and introducing the so-called effective fibre length

$$L_{\text{eff}} \triangleq \int_0^L e^{-\alpha z'}\, dz' = \frac{1-e^{-\alpha L}}{\alpha}, \tag{2.91}$$

the solution of (2.89) has the following simple closed-form expression

$$Q(z,t) = Q(0,t)\exp\left(\imath\gamma P_0 \cdot L_{\text{eff}}\right), \tag{2.92}$$

which can be interpreted as a NL phase shift. Note, for a long propagation distance L, *i.e.,* $\alpha L \gg 1$, we may use the following expression for asymptotic effective length $L_{\text{eff}}^{(a)} = 1/\alpha$. Typically, for SSMF with attenuation $\alpha = 0.2$ dB km at wavelength $\lambda_0 = 1550$ nm, the straight estimation of effective length gives $L_{\text{eff}} \approx 21$ km. Thus, using a long distance approximation, we have

$$P(t) \approx P_0(t)\exp\left(2\imath\,\phi_{\text{NL}}\right), \tag{2.93}$$

where $\phi_{\text{NL}} = (\gamma/\alpha)\, P_0$ denotes the phase-shift induced by the optical Kerr effect which depends on optical signal power. Eq. (2.93) shows that a signal phase is inherently corrupted due to optical Kerr nonlinearity in fibre, as a result, the spectrum of the signal broadens and this spectral broadening depends on the signal power. However, the NL effects in a fibre does not change the signal amplitude in the time domain.

2.2.8 Numerical solution: SSFT-method

Despite the fact that both the stochastic NLSE Eq. (2.46) and Manakov equation (ME) Eq. (2.52) describe light propagation through an optical fibre in a sufficiently accurate way, a general closed form solution including fibre loss, chromatic dispersion, fibre nonlinearity and amplifier noise does not exist. Nevertheless, the solution of the NLSE can be effectively approximated via a *pseudo-spectral* method, which is known as the split-step Fourier transform (SSFT) method [57]. This method has become prevalent owing to the implementation of the fast Fourier transform that significantly reduces

the complexity of the split-step algorithms, and enables a computationally tractable method to find a numerical solution of either NLSE or ME[8].

According to this method, decomposing Eq. (2.52) into its linear (dispersive) and NL (Kerr effect) constituents[9], *i.e.,*

$$\frac{\partial \mathbf{Q}(z,t)}{\partial z} = \left(\hat{\mathcal{D}} + \hat{\mathcal{N}}\right) \mathbf{Q}(z,t)\,, \tag{2.94}$$

the exact solution can be readily expressed via the exponential operator as follows

$$\mathbf{Q}(z,t) = \mathbf{Q}(0,t) \exp\left[\left(\hat{\mathcal{D}} + \hat{\mathcal{N}}\right) z\right], \tag{2.95}$$

where $\mathbf{Q}(0,t)$ stands for the input slowly-varying amplitude at $z = 0$, and the linear operator $\hat{\mathcal{D}}$ and NL operator $\hat{\mathcal{N}}$ are defined as

$$\hat{\mathcal{D}} \triangleq \imath \frac{\beta_2}{2} \frac{\partial^2}{\partial t^2}\,, \tag{2.96}$$

$$\hat{\mathcal{N}} \triangleq \imath \frac{8\gamma}{9} \mathbf{Q}^{\dagger}(z,t)\, \mathbf{Q}(z,t) - \frac{\alpha}{2}\,. \tag{2.97}$$

Thus, the so-called split-step type exact solution can be written as

$$\mathbf{Q}(z+\Delta,t) = \exp\left[\left(\hat{\mathcal{D}} + \hat{\mathcal{N}}\right)\Delta\right] \mathbf{Q}(z,t)\,, \tag{2.98}$$

where Δ is the step size.

One of the key ideas of SSFT is that the action of the sum of two operators are essentially equivalent to the sum of the actions of the two operators separately. If the step Δ is sufficiently small, we have

$$\mathbf{Q}(z+\Delta,t) \approx \exp\left(\hat{\mathcal{D}}\Delta\right) \exp\left(\hat{\mathcal{N}}\Delta\right) \mathbf{Q}(z,t)\,. \tag{2.99}$$

The operator on the right-hand side of Eq. (2.99) can be expanded by using the Baker-Campbell-Hausdorff formula for non-commutative operators [58], which gives an analytical insight into the accuracy on split-step approximated

[8]There are other methods to solve propagation equations available, such as the time-domain finite difference method, the inverse scattering transform (see [45] and refs therein).

[9]Here the dispersion slope as well the impact of inelastic scattering effects are omitted.

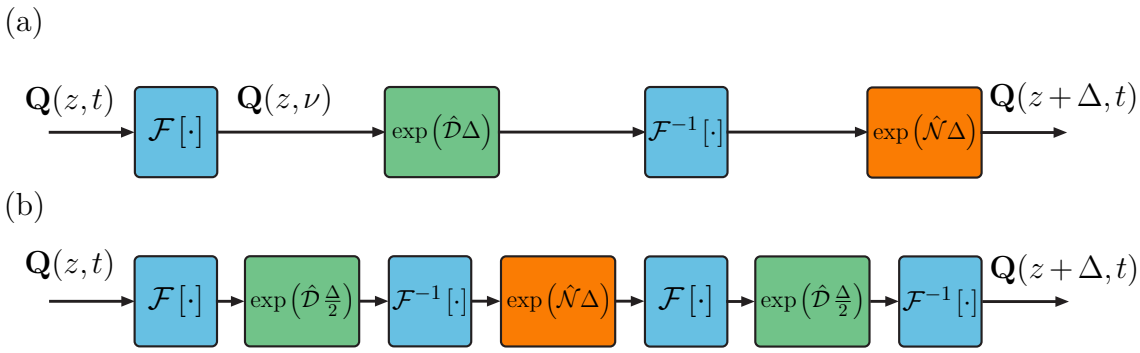

Figure 2.6: Schematic representation of a SSFT method with the bulk step (a) and symmetrised (b) schemes; $\mathcal{F}[\cdot]$ and $\mathcal{F}^{-1}[\cdot]$ stand for the fast forward and inverse Fourier transforms, respectively.

solutions

$$\exp\left(\hat{\mathcal{D}}\Delta\right)\exp\left(\hat{\mathcal{N}}\Delta\right)=\exp\left[\hat{\mathcal{D}}\Delta+\hat{\mathcal{N}}\Delta+\left[\hat{\mathcal{D}},\hat{\mathcal{N}}\right]\frac{\Delta^2}{2}+\left[\hat{\mathcal{D}}-\hat{\mathcal{N}},\left[\hat{\mathcal{D}},\hat{\mathcal{N}}\right]\right]\frac{\Delta^3}{12}+\ldots\right], \quad (2.100)$$

where $\left[\hat{\mathcal{D}},\hat{\mathcal{N}}\right]\triangleq\hat{\mathcal{D}}\hat{\mathcal{N}}-\hat{\mathcal{N}}\hat{\mathcal{D}}$ stands for the commutator. It can be seen that if $\Delta\to 0$, the solution given by Eq. (2.99) converges to the exact solution given by Eq. (2.98). Since the operators $\hat{\mathcal{D}}$ and $\hat{\mathcal{N}}$ are not commutative by nature, the accuracy of the SSFT based on the approximation Eq. (2.99) – the so-called *bulk* step method is of the second-order in the step size, *i.e.*, $O\left[\Delta^2\right]$.

In order to improve the accuracy of Eq. (2.99), we can resort to an alternative split-step scheme. This scheme can be realised by applying the dispersion operator into two equal parts before and after the NL operator [59]

$$\mathbf{Q}(z+\Delta,t)\approx\exp\left(\hat{\mathcal{D}}\frac{\Delta}{2}\right)\exp\left(\hat{\mathcal{N}}\Delta_{\text{eff}}\right)\exp\left(\hat{\mathcal{D}}\frac{\Delta}{2}\right)\mathbf{Q}(z,t), \quad (2.101)$$

where

$$\Delta_{\text{eff}}=\int_z^{z+\Delta}dz'\exp\left(-\alpha z'\right)=\frac{(1-\exp(-\alpha\Delta))\exp(-\alpha z)}{\alpha}. \quad (2.102)$$

The approximation Eq. (2.101) is recognised as the *symmetrised* SSFT. Similarly, the error term in this scheme can be found via the Baker-Campbell-

Hausdorff formula as follows

$$\exp\left(\hat{\mathcal{D}}\frac{\Delta}{2}\right)\exp\left(\hat{\mathcal{N}}\Delta\right)\exp\left(\hat{\mathcal{D}}\frac{\Delta}{2}\right)=\exp\left(\hat{\mathcal{D}}\Delta+\hat{\mathcal{N}}\Delta + \left[\hat{\mathcal{N}}+\frac{\hat{\mathcal{D}}}{2},\left[\hat{\mathcal{N}}+\frac{\hat{\mathcal{D}}}{2}\right]\right]\frac{\Delta^3}{6}+\ldots\right). \tag{2.103}$$

Eq. (2.103) indicates that the dominant error term is proportional to Δ^3, which provides much better accuracy in comparison with the bulk step case, Eq. (2.99). The block diagrams of the numerical implementation of the SSFT method for both (a) bulk and (b) symmetrised cases are illustrated in Fig. 2.6.

2.3 Generic fibre-optic communication system

In this section, we provide a brief description of the general structure of fibre-optic communication systems (a detailed description can be found in, *e.g.,* [60]). We also briefly introduce the numerical methods for estimating the key figure of merit (MI) that characterises the performance of modern optical communication systems (see, *e.g.,* [61, 62, 63, 64]). The main advantage of employing the MI as a main performance metric relies on its straightforward applicability to modern coded communication systems. The MI is chosen for use as a performance metric throughout this thesis.

2.3.1 Coherent fibre-optic communication system

Any fibre-optic communication system consists of a *transmitter*, a *fibre channel* (physical channel) and a *receiver*. Fig. 2.7 illustrates a general block-diagram of a coherent fibre-optic communication system. The transmitter realises a mapping between the input information bits $\boldsymbol{b}$ and the continuous-time input optical waveforms. During the propagation in a fibre, the signal waveforms are impaired by the linear additive noise arising from optical amplifiers. They also are broadened in the time domain due to the presence of the chromatic dispersion in a fibre, and distorted by Kerr NL effects. Moreover, the polarisation state of the signal waveforms are randomly changed, which stems from random imperfections in a realistic fibre.

Then, a receiver converts the output optical waveforms into electrical baseband signals via coherent detection, and eventually maps to output bits $\hat{\boldsymbol{b}}$. The DSP block at the receiver in Fig. 2.7 encodes the information bits into a digital signal providing forward error correction. DAC and ADC realise a conversion between the digital and the analog domain. An I-Q modulator is commonly a nested Mach-Zehnder modulator (MZM), which is able to modulate both

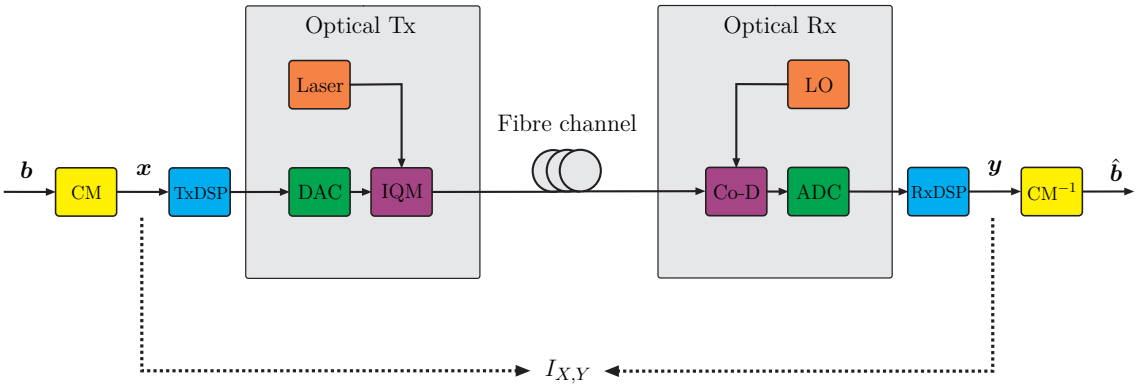

Figure 2.7: Schematic of a generic fibre-optic link with a coherent detection. Laser commonly implies a semiconductor laser. LO, IQM, Co-D stand for a local oscillator, an I-Q modulator, and a coherent detector (demodulator), respectively; TxDSP and RxDSP indicate the digital signal processing blocks at the transmitter (Tx) and at the receiver (Rx) sides, respectively; DAC and ADC stand for a digital-to-analog and a analog-to-digital converter, respectively. CM and CM^{-1} are the coded modulation and demodulation blocks, respectively (beyond the scope of this Thesis). $I_{X,Y}$ denotes the mutual information Eq. (2.107) per symbols in $\boldsymbol{x}$ / symbols out $\boldsymbol{y}$.

the signal amplitude and the phase covering a complex plane. Specifically, the I-Q modulator consisting of two parallel MZMs can be used to generate quadrature phase-shift keying (QPSK) signal modulation format. Multi-level driving signals can be used to change modulation format and/or pulse shape. Thus, these techniques allow the possibility to develop advanced multi-level signal modulation formats [65]. The propagation of the signal waveform over the optical fibre channel is comprehensively described in the previous Sec. 2.2. The received signal is coherently demodulated, acquiring the information of both amplitude and phase, and then, in a linear manner, transfers this into an electrical waveform. The coherent receiver usually includes a local oscillator, which is phase and frequency locked to the signal carrier[10]. In the case of non-linearity compensation (NLC) systems, the DSP block also includes either the electrical dispersion compensation (EDC) [67, 68] or the multi-channel digital back propagation (DBP) algorithm (*i.e.,* a numerical solution of NLSE/ME in digital domain in its inverse form by changing the sign of all the propagation parameters) [66]. Eventually, the output information bits $\hat{\boldsymbol{b}}$ are retrieved from the fully processed demodulated digital signal. The MI $I_{X,Y}$ between the output $\boldsymbol{y}$ and input $\boldsymbol{x}$ random symbols is also indicated in Fig. 2.7.

[10]The phase and frequency locking is, nowadays, achieved using DSP [66].

2.3.2 Mutual information computation

The MI has recently become a popular figure of merit for optical communication systems. In this section, we intend to clarify the numerical computation of the MI under the assumption of an additive white Gaussian noise (AWGN) channel. Such a quantity is highly important as it enables to obtain an information rate that can be achieved by a receiver based on Gaussian assumption [64]. It should also be noted that the MI, which is under that Gaussian memoryless channel assumption is essentially a lower bound on the MI of the channel with memory, usually found in the analysis of NL optical channel.

Let us assume a complex-input complex-output, discrete-time, memoryless channel characterised via the channel law PDF $p_{Y|X}(y|x)$. We also assume that the transmitted symbols belongs to a discrete signal constellation with $M = e^m$ points, where $m = \log M$ is the number of nats per symbol. The symbols are denoted by $x_i \in \mathbb{C}$, with $i = 1, 2, ..., M$, and $\mathbb{C}$ represents the set of complex numbers. We assume the symbols are equally likely, *i.e.,* they are transmitted with equal probabilities, and the PMF for is given by

$$p_X(x) \triangleq \mathbb{P}\left[X = x_i\right] = \frac{1}{M}. \tag{2.104}$$

The joint probability distribution $p_{X,Y}(x, y)$ (see, Bayes' theorem) [69] can be decomposed as

$$p_{XY}(x, y) = p_{Y|X}(y|x)\, p_X(x)\,. \tag{2.105}$$

Next, using the MI definition Eq. (2.14), for equally likely input symbols, the MI per polarisation can be expressed as follows

$$I_{X,Y} = \mathbb{E}\left[\log \frac{p_{Y|X}(Y|X)}{p_Y(Y)}\right], \tag{2.106}$$

using the law of total probability, we have

$$I_{X,Y} = \frac{1}{M} \sum_{i=1}^{M} \int_{\mathbb{C}} dy\, p_{Y|X}(y|x_i) \log \frac{p_{Y|X}(y|x_i)}{\frac{1}{M}\sum_{j=1}^{M} p_{Y|X}(y|x_j)}\,, \tag{2.107}$$

expressing the logarithm of a product as a sum of the logarithms, it yields

$$I_{X,Y} = m + \frac{1}{M} \sum_{i=1}^{M} \int_{\mathbb{C}} dy\, p_{Y|X}(y|x_i)\, g_i(y)\,, \tag{2.108}$$

where we have introduced

$$g_i(y) \triangleq \log \frac{p_{Y|X}(y|x_i)}{\frac{1}{M}\sum_{j=1}^{M} p_{Y|X}(y|x_j)}. \tag{2.109}$$

Obviously, this MI is a lower bound on the AWGN channel capacity due to the assumption of a discrete uniform input distribution.

Let us now consider a memoryless AWGN channel, *i.e.,*

$$Y = X + Z\,, \tag{2.110}$$

where Z is a complex, zero-mean, circularly symmetric Gaussian RV with total variance $\sigma_z^2 \triangleq \mathrm{Var}[Z] = \mathbb{E}\left[|Z|^2\right] = \mathbb{E}\left[|Y-X|^2\right]$, which represents an additive noise. The average symbol energy is assumed to be equal to $\sigma_x^2 = \mathbb{E}\left[|X|^2\right]$. Therefore, the signal-to-noise ratio (SNR) is defined by the following ratio

$$\mathrm{SNR} \triangleq \frac{\sigma_x^2}{\sigma_z^2} = \frac{\mathbb{E}\left[|X|^2\right]}{\mathbb{E}\left[|Y-X|^2\right]}\,. \tag{2.111}$$

The corresponding channel law of the AWGN channel is given by the Gaussian distribution, *i.e.,*

$$\begin{aligned} p_{Y|X}(y|x) &= p_Z(y-x) \\ &= \frac{1}{\pi\sigma_z^2}\exp\left(-\frac{|y-x|^2}{\sigma_z^2}\right). \end{aligned} \tag{2.112}$$

It is worth emphasising that the channel capacity[11] (see Eq. (2.24)) in nats per symbol of the AWGN Eq. (2.110) under averaged power constrained is well-known and is given by the famous *Shannon formula*, *i.e.,*

$$C = \log\left(1+\mathrm{SNR}\right), \tag{2.113}$$

where SNR is defined by Eq. (2.111).

[11]The maximum number of bits per symbol that can be reliably transmitted over an AWGN channel. The AWGN capacity can be achieved by applying the continuous Gaussian input distribution assuming the input-output relationship Eq. (2.110) and the Gaussian channel law Eq. (2.112).

Substituting Eq. (2.112) into Eq. (2.109), we obtain

$$g_i(y) = \log \frac{\exp\left(-\frac{|y-x_i|^2}{\sigma_z^2}\right)}{\sum_{j=1}^{M} \exp\left(-\frac{|y-x_i|^2}{\sigma_z^2}\right)} \tag{2.114}$$

$$= \log \frac{\exp\left(-\frac{|z_i(y)|^2}{\sigma_z^2}\right)}{\sum_{j=1}^{M} \exp\left(-\frac{\left|z_i(y)+d_{ij}\right|^2}{\sigma_z^2}\right)}, \tag{2.115}$$

where d_{ij} denotes the deference between two symbols x_i and x_j, and then, taking into account that $|z_1+z_2|^2 = |z_1|^2 + |z_2|^2 + 2\,\mathrm{Re}\,[z_1 \bar{z}_2]$, it reads

$$g_i(y) = -\log \sum_{j=1}^{M} \exp\left(-\frac{|d_{ij}|^2 + 2\,\mathrm{Re}\,[\bar{z}_i(y)\, d_{ij}]}{\sigma_z^2}\right), \tag{2.116}$$

where $\mathrm{Re}\,[\cdot]$ stands for a real part of complex-valued RV. Thus, using Eq. (2.116), the MI in Eq. (2.108) reads

$$I_{X,Y} = m - \frac{1}{M} \sum_{i=1}^{M} \int_{\mathbb{C}} dy\, p_{Y|X}(y|x_i) \log \sum_{j=1}^{M} \exp\left(-\frac{|d_{ij}|^2 + 2\,\mathrm{Re}\,[\bar{z}_i(y)\, d_{ij}]}{\sigma_z^2}\right). \tag{2.117}$$

Since Z_i is a circularly symmetric zero-mean complex Gaussian RV, the complex conjugate of $\bar{z}_i(y)$ can be dropped. Moreover, the sub-index i can be omitted as well, because the statistics of RV Z does not depend on i. Hence, changing the integration over y to z gives

$$I_{X,Y} = m - \frac{1}{M} \sum_{i=1}^{M} \int_{\mathbb{C}} dz\, p_Z(z) \log \sum_{j=1}^{M} \exp\left(-\frac{|d_{ij}|^2 + 2\,\mathrm{Re}\,[z\, d_{ij}]}{\sigma_z^2}\right). \tag{2.118}$$

Formally, the integral in Eq. (2.118) can be numerically computed via the following Monte-Carlo approximation, *i.e.,*

$$\int_{\mathbb{C}} dr\, p_R(r)\, g(r) \approx \frac{1}{N_s} \sum_{n=1}^{N_s} g\left(r^{(n)}\right) \tag{2.119}$$

where $R \in \mathbb{C}$ is the complex-valued RV, $g(r)$ is an arbitrary real-valued function, $r^{(n)}$, $n = 1, 2, ..., N_s$ are the random samples drawn from the distribution function $p_R(r)$, and N_s is the number of random samples.

Using Eq. (2.119), we have the following ready-to-use Monte-Carlo ap-

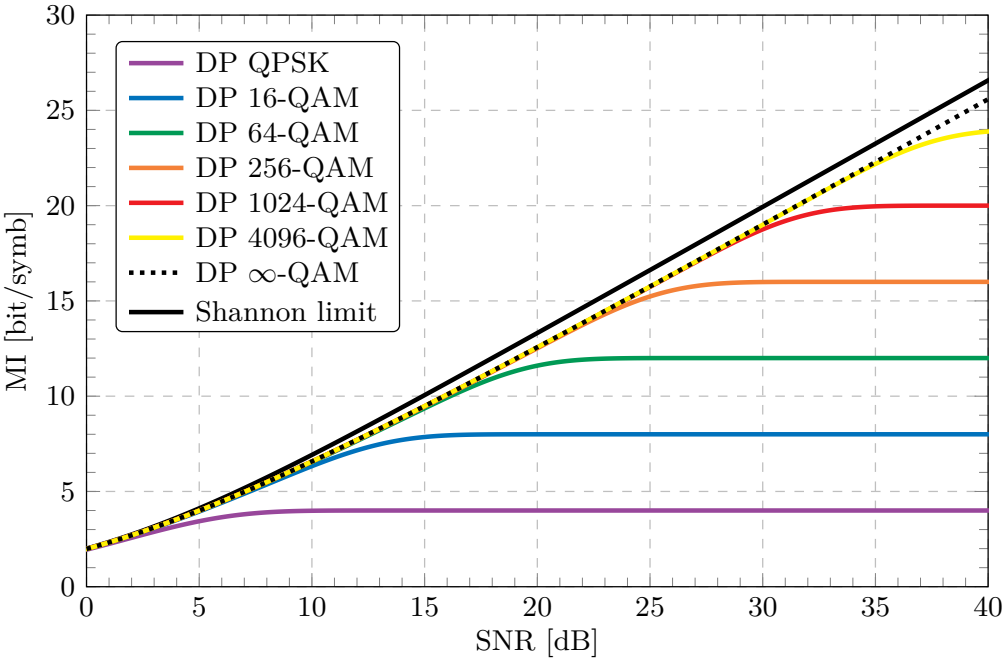

Figure 2.8: MI as a function of SNR for AWGN channel model for two polarisation states based on Eq. (2.125).

proximation for the integral in Eq, (2.118)

$$I_{X,Y} \approx m - \frac{1}{MN_s} \sum_{n=1}^{N_s} \sum_{i=1}^{M} \log \sum_{j=1}^{M} \exp\left(- \frac{|d_{ij}|^2 + 2\,\mathrm{Re}\,[z\, d_{ij}]}{\sigma_z^2} \right). \tag{2.120}$$

Alternatively, the integral in Eq. (2.118) can also be numerically calculated using Gauss-Hermite quadrature, *i.e.*, for any function $g(r)$ with $2L^{\text{th}}$ -order bounded derivative, the L-point Gauss-Hermite quadrature is given by

$$\int_{\mathbb{C}} dr \, \exp\left(-|r|^2\right) g(r) \approx \sum_{l_1=1}^{L} \alpha_{l_1} \sum_{l_2=1}^{L} \alpha_{l_1}\, g(\xi_{l_1} + \imath \xi_{l_2}), \tag{2.121}$$

where ξ_k are the k^{th} -root of the Hermite polynomial

$$H_L(x) \triangleq L! \sum_{r=0}^{\lfloor L/2 \rfloor} \frac{(-1)^r}{r!(L-2r)!} (2x)^{L-2r}, \tag{2.122}$$

and the weight functions are defined as

$$\alpha_l \triangleq \frac{\sqrt{\pi}\, 2^{L-1} L!}{\left[L H_{L-1}(\xi_l)\right]^2}. \tag{2.123}$$

If $L \to \infty$, the approximation in Eq. (2.121) would be exact. The values of α_l and ξ_l can be readily found numerically, which determines the trade-off between the speed of computation and the precision of quadrature approximation.

Substituting Eq. (2.112) into Eq. (2.118), and making the change of variables $z = \sigma_z r$, we obtain

$$I_{X,Y} = m - \frac{1}{M\pi} \sum_{i=1}^{M} \int_{\mathbb{C}} dr \, \exp\left(-|r|^2\right) \log \sum_{j=1}^{M} \exp\left(-\frac{|d_{ij}|^2 + 2\sigma_z \operatorname{Re}[r\, d_{ij}]}{\sigma_z^2}\right). \tag{2.124}$$

Applying Eq. (2.121) to Eq. (2.124), we obtained the following ready-to-use approximation of MI

$$I_{X,Y} \approx m - \frac{1}{M\pi} \sum_{i=1}^{M} \sum_{l_1=1}^{L} \alpha_{l_1} \sum_{l_2=1}^{L} \alpha_{l_2} \log \sum_{j=1}^{M} \exp\left(-\frac{|d_{ij}|^2 + 2\sigma_z \operatorname{Re}\left[(\xi_{l_1} + \imath\xi_{l_2})\, d_{ij}\right]}{\sigma_z^2}\right). \tag{2.125}$$

Let us now consider the continuous inform input distribution, which can be corresponded as a quadrature amplitude modulation (QAM) signalling with infinite cardinality, *i.e.,* "∞-QAM" signal modulation format. Therefore, the input $p_X(x)$ is given by the PDF

$$p_X(x) = \frac{1}{2\sqrt{3}\,\sigma_x}, \quad x \in \left[-\sigma_x\sqrt{3}, \sigma_x\sqrt{3}\right]. \tag{2.126}$$

Thus, the output of distribution $p_Y(y)$ in Eq. (2.10) of channel in Eq. (2.112) is then equal to

$$p_Y(y) = \frac{1}{4\sqrt{3}\,\sigma_x}\left[\operatorname{erf}\left(\frac{\sigma_x\sqrt{3}+y}{\sqrt{2}\,\sigma_z}\right) + \operatorname{erf}\left(\frac{\sigma_x\sqrt{3}-y}{\sqrt{2}\,\sigma_z}\right)\right], \tag{2.127}$$

where $\operatorname{erf}(\cdot)$ is the Gauss error function.

The conditional differential entropy Eq. (2.9) has the following closed form solution

$$h_{Y|X} = \int_{\mathcal{X}} dx\, p_X(x)\, h_{Y|X=x} \tag{2.128}$$

$$= \frac{1}{2}\log\left(2\pi e\,\sigma_z^2\right) \int_{\mathcal{X}} dx\, p_X(x) = \frac{1}{2}\log\left(2\pi e\,\sigma_z^2\right). \tag{2.129}$$

where we accounted the well-known entropy of the Gaussian distribution

$h_{Y|X=x}$, which is

$$h_{Y|X=x} \triangleq -\int_{\mathcal{Y}} dy\, p_{Y|X}(y|x) \log p_{Y|X}(y|x) \tag{2.130}$$

$$= \frac{1}{2} \log\left(2\pi e\, \sigma_z^2\right). \tag{2.131}$$

The output differential entropy h_Y can be approximated via Monte Carlo integration

$$h_Y \triangleq -\int_{\mathcal{Y}} dy\, p_Y(y) \log p_Y(y) \tag{2.132}$$

$$\gtrsim \frac{1}{N_s} \sum_{n=1}^{N_s} g\left(y_{(n)}\right), \tag{2.133}$$

where $g\left(y_{(n)}\right) \triangleq -\log\left[p_Y\left(y_{(n)}\right)\right]$ is the real-valued function, and the random samples $y_{(n)}$ are assumed to be drawn from the normal distribution with $n = 1, \ldots, N_s$, and N_s is the number of random samples. Thus, by using the decomposition Eq. (2.18) together with the differential entropies in Eqs (2.132) and (2.128), we can numerically evaluate the MI for the input QAM-signal with infinite cardinality.

The relationship between the MI for M-QAM with cardinalities up to $M \leq 4069 = 2^{12}$ (maximum spectral efficiency of 2×12 bits per symbol) and the SNR for the AWGN channel model is illustrated in Fig. 2.8, where the MI is numerically computed by using Eq. (2.125). The ∞-QAM case is also plotted. Note that, at high SNR level, the gap between the MI given by ∞-QAM input signal and the capacity of of the AWGN channel Eq. (2.113) is exactly equal to $6/(\pi e) \approx 1.53$ dB, which is often referred to as *ultimate shaping gain* (see, *e.g.,* [70, Sec. IV-B], [71, , Sec. VIII-A], [72, Sec. IV-B]).

2.3.3 Achievable information rate

Assuming the discrete complex input X with an arbitrary fixed PMF $P_X(x)$, and continuous output Y, the MI in Eq. (2.15), *i.e.,* the maximum achievable rate for a fixed memoryless channel can be expressed as follows

$$I_{X,Y} = \sum_{x \in \mathcal{X}} P_X(x) \int_{\mathbb{C}} dy\, p_{Y|X}(y|x) \log \frac{p_{Y|X}(y|x)}{\sum_{x' \in \mathcal{X}} p_{Y|X}(y|x') P_X(x')}, \tag{2.134}$$

where $\mathbb{C}$ represents the set of complex numbers. In order to compute Eq. (2.134) the knowledge of the channel statistics, *i.e.,* a channel law

$p_{Y|X}(y|x)$ is obligatory. In the general case of optical fibre channel, this function has no exact analytical expression. However, the MI can be lower-bounded as follows [73]

$$R_{X,Y} = \sum_{x\in\mathcal{X}} P_X(x) \int_{\mathbb{C}} dy\, p_{Y|X}(y|x) \log \frac{q_{Y|X}(y|x)}{\sum_{x'\in\mathcal{X}} q_{Y|X}(y|x')P_X(x')} \leq I_{X,Y} \leq C^{\text{mem}}, \tag{2.135}$$

where $q_{Y|X}(y|x)$ is the *auxiliary* channel law [74, 64], which is an approximation of the real channel $p_{Y|X}(y|x)$, and is commonly taken to be Gaussian distribution with noise variance σ^2, *i.e.,*

$$q_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-x)^2}{2\sigma^2}\right). \tag{2.136}$$

The information rate $R_{X,Y}$ in nats per symbol in Eq. (2.135) can be essentially achieved by using the optimal receiver for a channel given by $q_{Y|X}$, instead of using a suboptimal receiver for a *true* channel law $p_{Y|X}$. Therefore, for Nyquist-spaced WDM transmission system, the achievable information rate (AIR) in nats per second can be estimated as

$$\text{AIR} = N_{\text{pol}} N_{\text{ch}} R_S \cdot R_{X,Y}\,, \tag{2.137}$$

where N_{pol} is the number of polarisation states, N_{ch} is the number of Nyquist-space WDM channels, and R_S is the symbol rate.

Modelling the channel as the AWGN Eq. (2.110), the channel capacity in nats per second can be estimated as follows

$$C = N_{\text{pol}} N_{\text{ch}} R_S \log(1+\text{SNR})\,, \tag{2.138}$$

where the SNR denotes the so-called effective SNR and will be introduced and discussed in Sec. 3.2.

2.4 Summary

In this chapter, the essential mathematical tools for further theoretical investigations of coherent fibre-optic communication systems were reviewed. The discussed theoretical background are vital for thorough understanding of the results presented in this thesis. The master models, such as NLSE, coupled-mode NLSE, and ME, which govern the propagation of optical pulses through a fibre were briefly introduced. The efficient approach for solving these equations known as the SSFT method was presented. Also, this chapter

includes some implications of information theory in fibre-optic communications systems.

Chapter 3

Modelling of fibre-optic communication system

In this chapter, we review some recently developed heuristic mathematical approaches commonly used to analytically estimate the performance of modern fibre-optic communication systems. These models will be used to analyse and describe some fundamental limits occurring due to the NL interactions process between the information optical signal and the ASE noise.

3.1 Optical amplification schemes

Despite the fibre loss is being very low (as mentioned in the Introduction Sec. 1.1), the signal power eventually vanishes with propagation distance. In this subsection, we reviewed the mathematical modelling of different optical amplification schemes. These include lumped EDFA amplification as well as DRA techniques. In the presence of optical amplification, the NLSE Eq. (2.46) can be written as [75, Ch. 2]

$$\imath \frac{\partial Q(z,t)}{\partial z} + \imath \frac{\alpha - g(z)}{2} Q(z,t) - \frac{\beta_2}{2} \frac{\partial^2 Q(z,t)}{\partial t^2} + \gamma \left|Q(z,t)\right|^2 Q(z,t) = N(z,t), \tag{3.1}$$

where $g(z)$ denotes the amplification factor.

3.1.1 EDFA-amplification

In long-haul fibre links, lumped EDFAs provide broadband amplification of the transmitted optical signals. Lumped optical amplification implies a periodic use of amplifiers with a fixed period L_{s} of the fibre-optic link called the fibre *span* length. A schematic of a WDM fibre-optic transmission system of a single fibre span is illustrated in Fig. (3.1). For multi-span EDFA-amplified

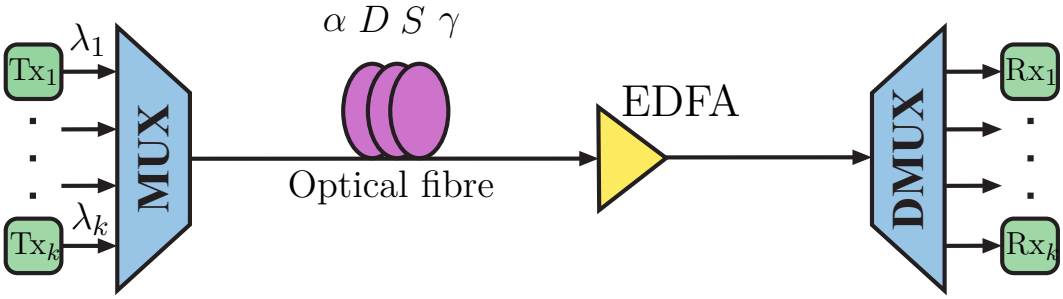

Figure 3.1: The scheme of a single fibre span WDM fibre-optic transmission system using a lumped EDFA

systems, the amplification factor $g(z)$ in Eq. (3.1) is given by

$$g(z) = \alpha L_{\mathrm{s}} \sum_{m=1}^{N_{\mathrm{s}}} \delta\left(z - mL_{\mathrm{s}}\right), \tag{3.2}$$

where L_{s} is the fibre span length, N_{s} is the total number of fibre spans in a link, $\delta(\cdot)$ is the Dirac delta-function. The autocorrelation function of the noise term $N(z,t)$ in Eq. (3.1) can be expressed as [76]

$$\mathbb{E}\left[N(z,t)\,N^*(z',t')\right] = \sigma_0^2\,\delta\left(t-t'\right)\sum_{m=1}^{N_{\mathrm{s}}}\delta\left(z-mL_{\mathrm{s}}\right), \tag{3.3}$$

where σ_0^2 is the PSD of ASE noise per unit bandwidth added per EDFA, *i.e.*,

$$\sigma_0^2 = n_{\mathrm{sp}}\left(G-1\right)\cdot h\nu_0, \tag{3.4}$$

where G is the EDFA gain, $h\nu_0$ is the average photon energy, ν_0 is the optical carrier frequency, h is the Planck's constant, and n_{sp} is the spontaneous-emission factor (the population-inversion factor) defined as [77]

$$n_{\mathrm{sp}} = \frac{N_1}{N_2 - N_1} \geq 1, \tag{3.5}$$

where N_1 and N_2 are the atomic populations for the ground and excited states, respectively. Obviously, the effect of spontaneous emission always adds certain fluctuations to the amplified signal. For periodically-spaced discrete EDFAs, the PSD S_{ASE} of linear ASE noise induced by the EDFA amplifier in each span of the link is nearly constant (white noise) and can be given by the well-known

expression [78, 79]

$$S_{\mathrm{ASE}}(\nu) = N_{\mathrm{pol}}\, n_{\mathrm{sp}} \Big[G(\nu) - 1 \Big] \cdot h\nu_0 \,, \tag{3.6}$$

where N_{pol} is the number of polarisation states, $G(\nu)$ is the gain provided by the EDFA amplification, which is commonly assumed to be precisely equal to the single fibre span loss, *i.e.,* $G = \exp(\alpha L_{\mathrm{s}})$, where α is the fibre attenuation (loss) parameter given in linear scale, L_{s} is the fibre span length.

Thus, the total ASE noise power can be given by integration of Eq. (3.6) over the total optical bandwidth B, giving

$$\begin{aligned} \sigma^2_{\mathrm{ASE}} &= \int_{-B/2}^{B/2} d\nu\, S_{\mathrm{ASE}}(\nu)\, H(\nu) \\ &= N_{\mathrm{pol}}\, n_{\mathrm{sp}} \cdot h\nu_0 \int_{-B/2}^{B/2} d\nu \Big[G(\nu) - 1 \Big] H(\nu) \,, \end{aligned} \tag{3.7}$$

where $H(\nu)$ denotes the response function of the applied matched filter used at the receiver usually corresponding to the worst case scenario. Because the ASE noise is inherently a linear noise, therefore, the noise accumulation occurs in a linear manner with transmission distance.

Assuming optical gain G provided by each EDFA to be constant within the reference bandwidth B_{ref}, thus, the total ASE noise power per fibre span is

$$\sigma^2_{\mathrm{ASE}} = N_{\mathrm{pol}}\, n_{\mathrm{sp}}\, (G-1) \cdot h\nu_0 \cdot B_{\mathrm{ref}} \,. \tag{3.8}$$

The factor n_{sp} is, in turn, directly related to the EDFA noise figure (NF) as follows [45]

$$\mathrm{NF} = 2n_{\mathrm{sp}} \left(1 - \frac{1}{G} \right) + \frac{1}{G} \,, \tag{3.9}$$

for relatively high gain $G \gtrsim 10$ dB (which is usually always the case), it can be effectively approximated as follows

$$n_{\mathrm{sp}} \approx \frac{1}{2} \mathrm{NF} \,. \tag{3.10}$$

Thus, Eq. (3.8) can be re-written as

$$\sigma^2_{\mathrm{ASE}} \approx \frac{1}{2} N_{\mathrm{pol}}\, (G-1) \cdot \mathrm{NF} \cdot h\nu_0 \cdot B_{\mathrm{ref}} \,, \tag{3.11}$$

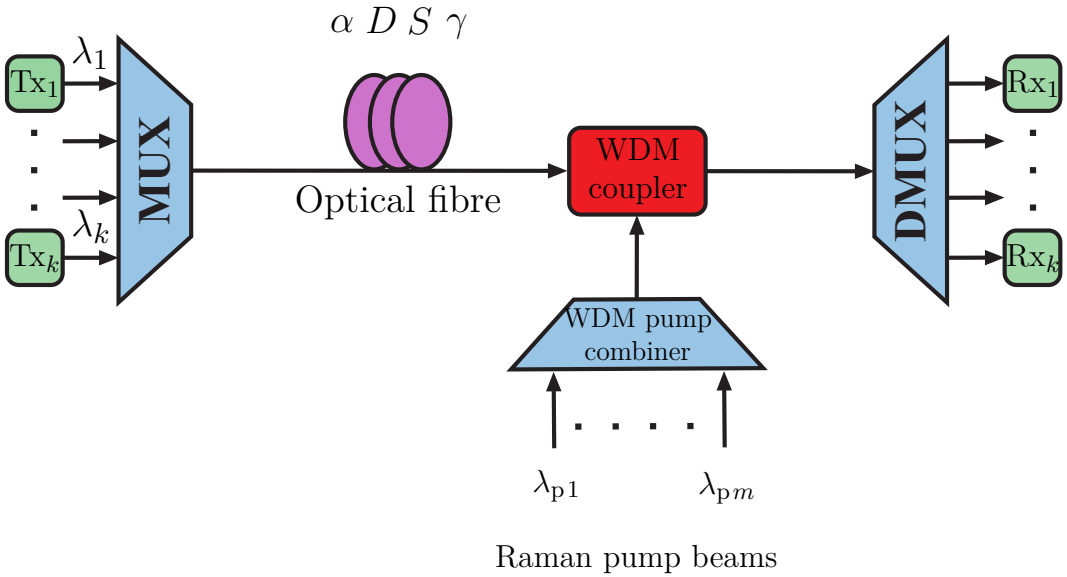

Figure 3.2: The scheme of a single fibre span WDM transmission system using the backward-pump geometry of DRA.

Eq. (3.11) is most commonly used since NF can be experimentally measured. The range of typical NF values for a practical EDFA is 4–7 dB.

3.1.2 Distributed Raman amplification

The nature of optical Raman amplification is essentially related to a phenomena called inelastic scattering (in particular, stimulated Raman scattering) between an incident light and molecular vibrations in a fibre. The scattered light (Stokes radiation) always has lower frequencies compared to incident radiation, *i.e.,* it scatters the energy due to this process. In this subsection, we examine the backward-pumped geometry of optical DRA (see Fig. 3.2). Omitting details, the process of DRA for co-polarised signal and pump beams obeys the following two coupled ordinary differential equations

$$\begin{cases} \dfrac{dn}{dz} = \gamma_R n_{\mathrm{p}} (n_{\mathrm{s}} + 1) - \alpha n \,, \\ \dfrac{dn_{\mathrm{p}}}{dz} = \gamma_R \, n_{\mathrm{p}} (n_{\mathrm{s}} + 1) - \alpha_{\mathrm{p}} n_{\mathrm{p}} \,, \end{cases} \tag{3.12}$$

where n and n_{p} are the photon numbers of the forward-signal and backward-pump beams, respectively; α and α_{p} are the absorption coefficients of the signal and pump photons, respectively; γ_R is the microscopic Raman gain coefficient. Unfortunately, the system (3.12) in its general form has no closed-form solution. However, it is possible to easily solve such systems numerically

by using, for example, the Runge-Kutta method [80].

Under the "small-signal" assumption, *i.e.,* the photons of the pump beam are assumed to be not depleted by signal photons during the propagation, and thus, the system (3.12) can be solved analytically. Hence, from the second equation, the number of the backward-pump photons n_{p} yields

$$n_{\mathrm{p}}(z) = n_{\mathrm{p}}(L)\exp\left[\alpha_{\mathrm{p}}\,(z-L)\right], \tag{3.13}$$

where $n_{\mathrm{p}}(L)$ denotes the input pump photon number at the distance $z = L$. Substituting Eq. (3.13) in the first equation of the system (3.12), we end up with the following non-homogeneous linear ordinary differential equation of order one for the forward-signal photon numbers

$$\frac{dn(z)}{dz} + \left[\alpha - \gamma_R\, n_{\mathrm{p}}(z)\right] n(z) = \gamma_R\, n_{\mathrm{p}}(z)\,. \tag{3.14}$$

It is obvious that Eq. (3.14) can be integrated straightforwardly via the method of variation of parameters (Lagrange method). The solution can then be expressed as

$$n(L) = n(0)\,G_{\mathrm{R}} + n_{\mathrm{sp}}\,, \tag{3.15}$$

where the introduced optical gain G_{DRA} provided by DRA is given by

$$G_{\mathrm{R}} = \exp\left[\alpha L - \gamma_R\, n_{\mathrm{p}}(L)\cdot L_{\mathrm{eff,p}}\right], \tag{3.16}$$

where $L_{\mathrm{eff,p}}$ is the effective fibre length for a pump beam, *i.e.,*

$$L_{\mathrm{eff,p}} \triangleq \frac{1-e^{-\alpha_{\mathrm{p}}L}}{\alpha_{\mathrm{p}}}\,, \tag{3.17}$$

and n_{sp} in (3.15) stands for the spontaneous emission noise photon number, which can be expressed in terms of optical launched power P_{p} of the backward Raman pump beam as follows [81]

$$\begin{aligned} n_{\mathrm{sp}} = \left(\frac{\alpha_{\mathrm{p}}}{g_R P_{\mathrm{p}}}\right)^{\frac{\alpha}{\alpha_{\mathrm{p}}}} \Bigg[\Gamma\left(1+\frac{\alpha}{\alpha_{\mathrm{p}}}, \frac{g_R P_{\mathrm{p}}}{\alpha_{\mathrm{p}}}\, e^{-\alpha_{\mathrm{p}} L_{\mathrm{s}}}\right) \\ - \Gamma\left(1+\frac{\alpha}{\alpha_{\mathrm{p}}}, \frac{g_R P_{\mathrm{p}}}{\alpha_{\mathrm{p}}}\right)\Bigg] \cdot \exp\left(\frac{g_R P_{\mathrm{p}}}{\alpha_{\mathrm{p}}}\right), \end{aligned} \tag{3.18}$$

where $\Gamma(a,\cdot)$ denotes the upper incomplete Gamma function with a complex parameter a, g_R is the modal peak Raman gain coefficient measured in $\left[\mathrm{W}^{-1}\,\mathrm{km}^{-1}\right]$. Assuming the Raman pumping entirely compensates the fibre

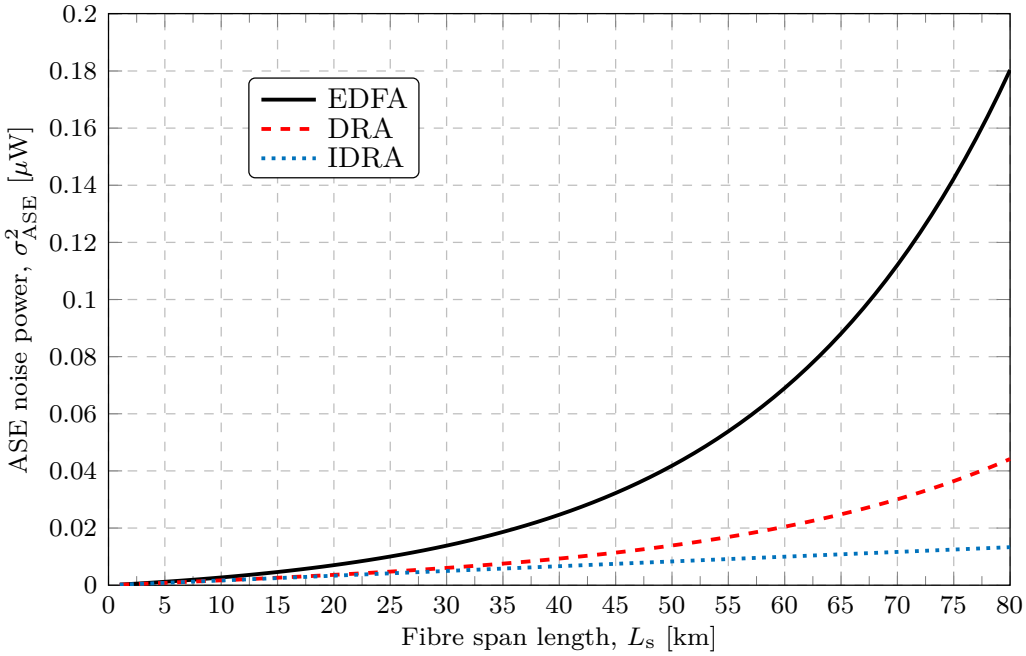

Figure 3.3: The ASE noise power against the fibre span length for different amplification schemes: EDFA Eq. (3.11), DRA Eq. (3.19) and IDRA Eq. (3.21).

loss, *i.e.*, $G_{\mathrm{R}} \approx 1$, then, from Eq. (3.16), the product $g_R P_{\mathrm{p}} \approx \alpha L / L_{\mathrm{eff,p}}$.

Thus, the total variance of the ASE noise per fibre span over the reference bandwidth B_{ref} due to optical counter-pumped fibre Raman amplification is defined as

$$\sigma^2_{\mathrm{ASE}} = N_{\mathrm{pol}} K_T n_{\mathrm{sp}} \cdot h\nu_0 \cdot B_{\mathrm{ref}}\,, \tag{3.19}$$

where K_T denotes the temperature dependent phonon occupancy factor, which is given by [82]

$$K_T = 1 + \frac{1}{\exp\left(\frac{h\Delta\nu}{k_B T}\right) - 1}\,, \tag{3.20}$$

where $\Delta\nu = \nu_{\mathrm{p}} - \nu_0$ denotes the frequency shift between the Raman pump frequency ν_{p} providing the distributed gain and the central frequency ν_0 of signal beam, k_B is the Boltzmann constant, T is the temperature of the fibre measured in [K]. For Raman amplified fibre-optical communication systems the value of K_T is typically between 1.1 and 1.3. Note that, since phonons are bosons, the second term in Eq. (3.20) appropriately represents the Bose-Einstein statistics. For ideal distributed Raman amplifier (IDRA), the total ASE variance per fibre span is

$$\sigma^2_{\mathrm{ASE}} = N_{\mathrm{pol}} K_T \alpha L \cdot h\nu_0 \cdot B_{\mathrm{ref}}\,, \tag{3.21}$$

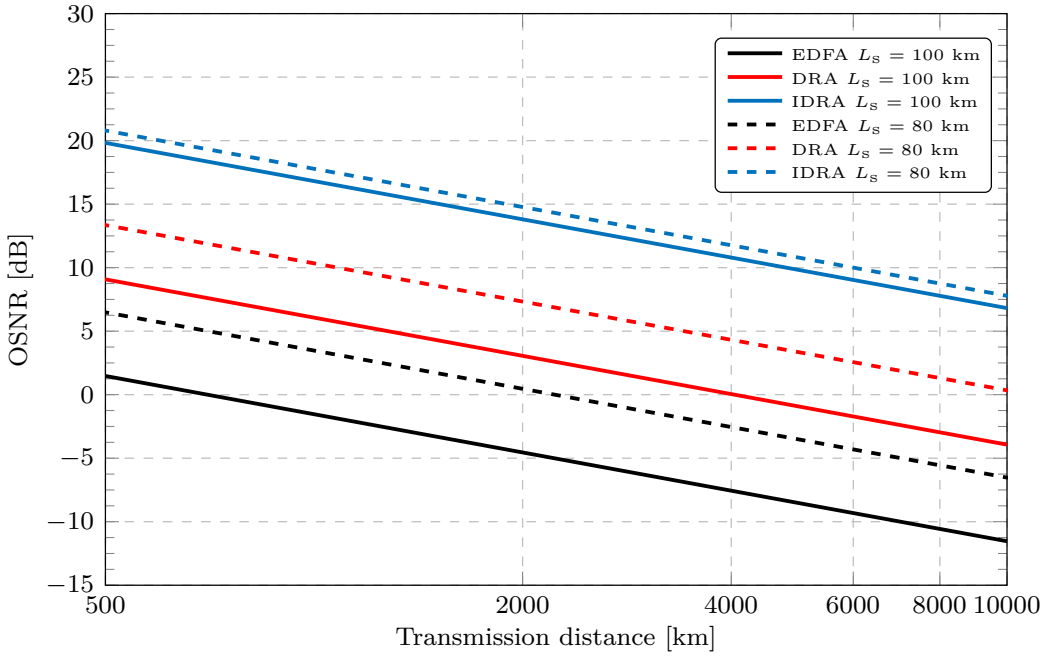

Figure 3.4: OSNR as a function of transmitted distance for two amplifier spacings: 80 km and 100 km using different amplifier schemes: EDFA with NF = 4.5 dB, DRA, and IDRA, assuming the fixed launched pump power 1 mW.

where L stands for the total length of the fibre link. The amplification factor $g(z) = \alpha$. Note that, the lumped amplification asymptotically tends to the distributed Raman amplification when $L_s \to 0$ that makes a perfect physical sense.

Fig. 3.3 shows how fast ASE noise grows with fibre span length using different amplification schemes, such as lumped EDFA, DRA and IDRA. The ASE noise EDFA displays rapid exponential grows, whilst ASE due to DRA growth much slower. IDRA grows linearly with fibre span length. In contract to the distributed amplification, the lumped amplification adds much more ASE noise for a fixed span length. This difference increases with the amplifier spacing.

Neglecting the influence of NL effects in a fibre, the linear ASE noise impairs the so-called optical signal-to-noise ratio (OSNR), which commonly is defined as follows

$$\mathrm{OSNR}|_{B_{\mathrm{ref}}=0.1\,[\mathrm{nm}]} \triangleq \frac{N_{\mathrm{pol}}\, R_S}{2B_{\mathrm{ref}}}\,\mathrm{SNR}\Big|_{B_{\mathrm{ref}}=0.1\,[\mathrm{nm}]}, \tag{3.22}$$

where $B_{\mathrm{ref}} = 0.1$ nm indicates the reference bandwidth, which is approximately

taken to be equal to 12.48 GHz that corresponds to a 0.1 nm resolution bandwidth of optical spectrum analysers at 1550 nm optical carrier wavelength (193.4 THz carrier frequency). Fig. 3.4 illustrates the OSNR degradations with transmission distance. A shorter amplifier spacing demands a higher OSNR at the receiver. Evidently, to come up with a longer span amplifier spacing, the launched signal power needs to be increased. This is doable and naively easy provided that the NL power-dependent Kerr effect is not under consideration. It should be emphasised that, in the purely linear regime, the distributed amplification schemes demonstrate better performance than the EDFA for a fibre-optic system with identical parameters.

3.2 Fibre-optic system analytical models

In this section, the existing analytical models for dispersion unmanaged coherent fibre-optic communication systems are review. The key idea of modelling such systems is that the joint effect of weak nonlinearity-induced distortions and relatively large accumulated dispersion can be effectively treated as an additive noise. This nonlinearity-induced noise can be relatively accurately approximated as a Gaussian noise process with zero-mean, and is commonly known as nonlinear interference (NLI). The NLI noise is always assumed to be statistically independent of the linear noise due to the ASE process, *i.e.,* NLI incoherently combines with ASE noise. The statistics of the power of the emerging NLI noise cubically grows with the transmitted power. It should also be emphasised that these noise-like distortions can be treated as a white Gaussian noise only in fibre-optic systems with no in-line chromatic dispersion compensation, *i.e.,* dispersion is assumed to be compensated electronically via digital signal processing at the receiver. Hence, these models can be generally represented via a discrete-time AWGN model

$$\mathbf{Y}_k = \mathbf{X}_k + \mathbf{Z}_k, \quad k \in \mathbb{Z} \tag{3.23}$$

where k is the time index, $\mathbf{Y}_k$ is the sequence of output symbols. The transmitted symbols $\mathbf{X}_k$ are assumed to be IID, having the same average transmitted power $P \triangleq \mathbb{E}\left[\mathbf{X}\mathbf{X}^{\dagger}\right]$, which is a crucial assumption since the optical Kerr effect has an instantaneous nature. Each symbol in the sequence $\mathbf{Z}_k$ represents circularly symmetric complex-valued white Gaussian RV and statistically independent of input symbols $\mathbf{X}_k$, *i.e.,* $\mathbf{Z}_k \sim \mathcal{CN}\left(0, \sigma^2_{\text{eff}}\right)$, where the total noise variance $\sigma^2_{\text{eff}} \triangleq \mathbb{E}\left[\mathbf{Z}\mathbf{Z}^{\dagger}\right]$ consists of the contribution of both linear ASE noise and NLI noise. Notably, the NLI noise circularity assumption fails at low

symbol rate regime.

3.2.1 Literature review

In recent years, the simple phenomenological analytical models that allow predictions of fibre-optic system performance degradation due to the optical Kerr effect have gained considerable popularity. Most of these mathematical models are based on the approximate solution of propagation equations (either the NLSE or ME) in the framework of the first-order perturbation analysis [83, 84, 85, 86]. In order to rapidly predict the performance of point-to-point links [87, 27, 28] as well as to enable efficient system design [88], such analytical models became of paramount importance since they provide a possibility to avoid the computational complexity associated with a numerical solution of the propagation equations. Particularly, this becomes significantly relevant especially in the case of ultra-wideband communications (*e.g.,* up to 1 THz signal bandwidth and beyond).

The first successful attempts to derive analytical closed-form expressions for estimating the efficacy and the performance of dispersion-uncompensated multi-span WDM communication systems taking into account optical Kerr nonlinearity have been made by Splett *et al.* [89] in 1993. In particular, in [89], assuming an ideal optical amplification, the *optimal* SNR and the resulting system capacity under the assumption of an AWGN channel (*i.e.,* assuming Gaussian channel law as well as implicitly considering nonlinearity as an additive Gaussian noise) as a function of, for instance, the bandwidth (BW) have been assessed. This means that the peaky behaviour of system performance due to the NL interference was tacitly presented. In 2001, the peaky non-monotonic behaviour of a lower bound on the information capacity has been shown by Mitra and Stark [90]. In their approach, in contrast to Splett approach, the Kerr nonlinearity was modelled as a multiplicative noise. For multi-user (*i.e.,* WDM) case, the NL interference, owing to the cross-phase modulation (XPM) effect was argued to be dominant. In the absence of both self-phase modulation (SPM) and FWM contributions, the XPM term in the frequency domain was modelled as a potential of Gaussian distribution with constant first and second order moments, linearising the NLSE thereby. The resulting linearised inhomogeneous NLSE was solved via a method of Green's functions, which is widely used in quantum mechanics (details can be found in [91, 92]). It should also be noted that derived closed-form expressions to estimate the instantaneous power of NL distortions in both Splett and Mitra-Stark models are not applicable directly for practical systems and remain quantita-

tively questionable since they were implicitly derived under the assumption of a flat power profile, *i.e.,* ideal optical amplification. However, these two models gave a clear idea about the qualitative communication system behaviour, as well as the existence of a power threshold as a consequence of optical Kerr effect. In 2002, Tang [93] considered a multi-span dense WDM system for both dispersion and dispersion-free fibres, and the peaky system performance behaviour was demonstrated. In 2003, the proposed analytical model for a dense WDM transmission with various system parameters has been numerically verified by Louchet [94]. In 2010, closed-form analytical expressions were derived by Chen and Shieh [95] for a densely-spaced coherent orthogonal frequency-division multiplexing (OFDM) signal.

Later on, similar analytical results were independently obtained and developed by Poggiolini and Carena [96, 97, 98], and the model has eventually become widely recognised as the Gaussian noise (GN) model. Johannisson, Karlson and Agrell have also independently published detailed analytical derivations considering wide signal spectrum in a strongly dispersive optical communication system [99, 100]. It is significant to note that this GN model remains inherently perturbative, *i.e.,* it assumes the smallness of the generated noise-like interference. This so-called NL interference occurs as a summation of all the FWM products induced by combinations of optical signal spectral components across its transmitted spectrum. The crucial assumption of the GN model is that the input optical signal is a Gaussian random process, which, in the frequency domain, can be represented as a periodic white Gaussian noise process of period $1/\nu_0$, with ν_0 standing for the centre frequency; *i.e.,* a grid of Dirac delta functions, and each frequency component is given by statistically independent Gaussian RVs. Mathematically, using the Karhunen-Loève expansion [101, Sec. 2.8-2], it yields the following initial condition for the electric field envelope $\mathbf{Q}(\nu)$ in the ME Eq. (2.52) in the frequency domain

$$\mathbf{Q}(\nu) = \sqrt{\nu_0} \sum_{m=-\infty}^{\infty} \boldsymbol{\zeta}_m \, \delta\left(\nu - m\nu_0\right), \tag{3.24}$$

where $\boldsymbol{\zeta}_m$ are the realisations of independent complex vector RVs $\boldsymbol{Z}_m \sim \mathcal{N}(0,1)$. The underlying factor of this assumption is that, during the propagation, the *Gaussianity* of the optical signal is supposed to be maintained by large accumulated chromatic dispersion in the link as well as the denser modulation format [102].

The main disadvantage of the conventional GN model is that it is unable to predict a signal modulation format dependence [20, 103, 61, 102] as well

as symbol rate dependence [104, 105] and the impact of memory effects [106]. Importantly, the first-order perturbation analysis is equivalent to the time-domain Volterra serious approach for the NL fibre channel. In 2012, Mecozzi and Essiambre evaluated the third-order Volterra series term [20], which is essentially corresponding to the first-order regular perturbation approximation of the NL distortion coefficient. Using the approach in [20], a more accurate input signal dependent analytical model was derived and validated by Dar [102, 107, 108, 109]. The modulation format dependent GN requires numerical integrations, for instance, the Monte-Carlo sampling [110]. However, it was principally assumed that the XPM has a dominant impact on the performance degradation, whilst Carena *et al.* took into account all NL contributions, such as SPM, XPM and FWM [111]. An alternative approach also relying on Mecozzi's seminal results was proposed by Sarena and Bononi, where the time-domain autocorrelation function of the NL cross-channel distortions has been examined and input signal distributions accounted for [112, 113]. However, this approach does not provide any ready-to-use closed form solutions, and so the numerical computations are compulsory.

3.2.2 Conventional Gaussian noise modelling

The performance of the WDM transmission systems can be estimated by introducing an *effective* SNR that includes both ASE noise and optical Kerr effect contributions. After coherent detection and EDC[1]

$$\mathrm{SNR} = \frac{P}{\sigma^2_{\mathrm{eff}}} \approx \frac{P}{N_{\mathrm{s}}\,\sigma^2_{\mathrm{ASE}} + P_{s-s}}\,, \tag{3.25}$$

where P is the average optical power per channel, σ^2_{ASE} is the overall ASE noise power, arising from EDFAs at the end of each span in a link, N_{s} is the total number of fiber spans in a link. The contribution of the deterministic S-S interactions is described as

$$P_{s-s} = N_{\mathrm{s}}^{\epsilon+1}\,\eta P^3\,, \tag{3.26}$$

where η is the NL distortion coefficient that quantifies the NL interference of one fibre span, and $\epsilon \in [0,1]$ denotes the coherence factor, which essentially characterises the decorrelation of the NL distortions between each fibre span in a link.

[1]The signal-ASE noise interaction is omitted in this subsection, and is considered later. Note that, the contribution of signal-ASE noise interaction is negligible in the presence of signal-signal NL interference.

The NL distortion coefficient η corresponding to the centre channel can be computed by the following double integral

$$\eta = \frac{16\gamma^2}{27R_S^2} \int_{-B/2}^{B/2} \int_{-B/2}^{B/2} d\nu_1 d\nu_2 \, \mathrm{rect}\left(\frac{\nu_1+\nu_2}{B}\right) |\,\mu(\nu_1,\nu_2)\,|^2 \,, \tag{3.27}$$

with the total WDM transmission bandwidth being $B \triangleq N_{\mathrm{ch}} \cdot \Delta\nu$, and $\Delta\nu$ denoting the channel spacing, R_S the symbol rate, and γ the fibre nonlinearity parameter, and rect$(\cdot)$ the rectangular function. The kernel function can be factorised as follows

$$\mu(\nu_1,\nu_2) = \rho(\nu_1,\nu_2)\,\phi(\nu_1,\nu_2) \,, \tag{3.28}$$

where the so-called *phased-array* factor $\phi(\nu_1,\nu_2)$ takes into account the accumulation of NL interaction process over the multi-span transmission system and can be expressed as

$$\phi(\nu_1,\nu_2) = \frac{1-\exp(\imath\,\Delta\beta(\nu_1,\nu_2)N_{\mathrm{s}}L_{\mathrm{s}})}{1-\exp(\imath\,\Delta\beta(\nu_1,\nu_2)L_{\mathrm{s}})} \,, \tag{3.29}$$

and the FWM efficiently factor yields

$$\rho(\nu_1,\nu_2) = \int_0^{L_{\mathrm{s}}} dz\, P(z) \exp(\imath\,\Delta\beta(\nu_1,\nu_2)\,z) \,, \tag{3.30}$$

where $P(z)$ defines the signal power profile along the fibre span, which depends on the applied optical amplification scheme, and $\Delta\beta$ is the FWM phase-mismatch, which can be expressed as follows[2] (see, *e.g.,* [114])

$$\Delta\beta(\nu_1,\nu_2) = \beta(\nu_1+\nu_2) - \beta(\nu_1) - \beta(\nu_2) + \beta(\nu_0) \tag{3.31}$$

$$\approx 4\pi^2\,|\beta_2|\,\nu_1\nu_2 \,, \tag{3.32}$$

where $\beta(\cdot)$ denotes the propagation constant as a function of frequency, and β_2 represents the second-order dispersion coefficient. For lumped EDFA, the signal power in Eq. (3.30) exponentially decays with distance, *i.e.,* $P(z) = e^{-\alpha z}$, therefore, there exists the closed form solution

$$\rho(\nu_1,\nu_2) = \frac{1-\exp\left[-\left(\alpha+\imath\,\Delta\beta(\nu_1,\nu_2)\right)L_{\mathrm{s}}\right]}{\alpha-\imath\,\Delta\beta(\nu_1,\nu_2)} \,. \tag{3.33}$$

[2]The central frequency ν_0 in Eq. (3.31) corresponds to zero-frequency.

In the case of non-ideal backward-pumped DRA, the signal power profile $P(z)$ can be defined as a solution of a system of two coupled differential equations, which govern the Raman process for a single co-polarised pump wavelength and signal wavelength travelling in the backward direction Eqs (3.12). In terms of power Eqs (3.12) can be readily re-written as (see, *e.g.,* [115, Eqs (1-2)])

$$\begin{cases} \dfrac{dP}{dz} = -\alpha P + g_R P P_{\mathrm{p}} \\ -\dfrac{dP_{\mathrm{p}}}{dz} = -\alpha_{\mathrm{p}} P_{\mathrm{p}} + \left(\dfrac{\nu_{\mathrm{p}}}{\nu}\right) g_R P_{\mathrm{p}} P, \end{cases} \tag{3.34}$$

where P_{p} is the pump power, g_R is the Raman gain coefficient[3] measured in $\left[\mathrm{W}^{-1}\,\mathrm{km}^{-1}\right]$, which is normalised by the affective area of the modes in the fibre, ν_{p} is the pump frequency. Neglecting the pump depletion term (*i.e.,* a term with a pre-factor (ν_{p}/ν)), the system Eqs (3.34) can be solved analytically. The resulting power profile for a backward-pumped DRA can be expressed as

$$P(z) = \exp\left\{\frac{g_R P_{\mathrm{p}}\left[\exp\left(-\alpha_{\mathrm{p}} z\right) - 1\right]}{\alpha_{\mathrm{p}}}\right\} \exp\left(-\alpha z\right). \tag{3.35}$$

Substituting Eq. (3.35) into the integral Eq. (3.30), we have the following closed form solution

$$\rho(\nu_1, \nu_2) = \frac{\Xi(\nu_1, \nu_2)}{\alpha_{\mathrm{p}}} \left(-\frac{g_R P_{\mathrm{p}}}{\alpha_{\mathrm{p}}}\right)^{\frac{\alpha - \imath\,\Delta\beta(\nu_1,\nu_2)}{\alpha_{\mathrm{p}}}} \exp\left(-\frac{g_R P_{\mathrm{p}}}{\alpha_{\mathrm{p}}}\right), \tag{3.36}$$

where the following function has been introduced

$$\begin{aligned} \Xi(\nu_1, \nu_2) \triangleq\ & \Gamma\left(-\frac{\alpha - \imath\,\Delta\beta(\nu_1,\nu_2)}{\alpha_{\mathrm{p}}}, -\frac{g_R P_{\mathrm{p}}}{\alpha_{\mathrm{p}}} \exp\left(\alpha_{\mathrm{p}} L_{\mathrm{s}}\right)\right) \\ & - \Gamma\left(-\frac{\alpha - \imath\,\Delta\beta(\nu_1,\nu_2)}{\alpha_{\mathrm{p}}}, -\frac{g_R P_{\mathrm{p}}}{\alpha_{\mathrm{p}}}\right), \end{aligned} \tag{3.37}$$

where $\Gamma(a, \cdot)$ denotes the upper incomplete Gamma function with a complex parameter a.

The above-mentioned coherence factor ϵ can be numerically evaluated as follows

$$\epsilon = \frac{1}{\ln 100} \ln\left(\frac{\eta|_{N_{\mathrm{s}}=100}}{\eta|_{N_{\mathrm{s}}=1}}\right) - 1. \tag{3.38}$$

[3]Hereinafter a depolarised pump source and no polarisation dependence of the Raman gain are assumed. In reality, the Raman gain is polarisation dependent; however, high gain efficiencies can still be realised with the method described in [116]

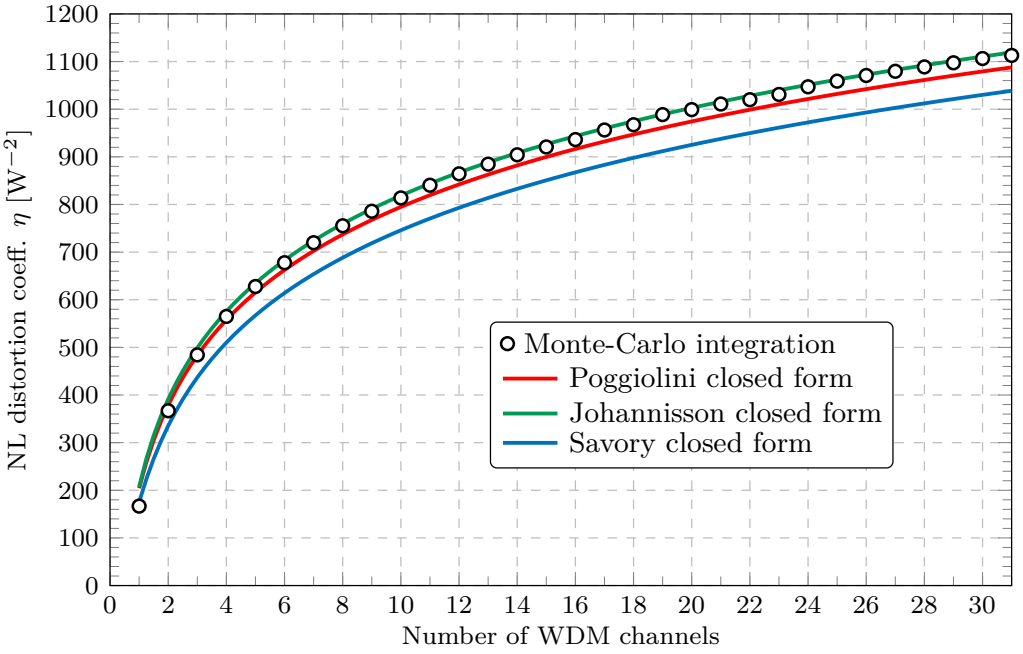

Figure 3.5: The NL distortion coefficient η as a function of number of ideal Nyquist-spaced WDM channels (*i.e.,* transmitted bandwidth). The symbol rate $R_S = 32$ GBaud. Black dots correspond to the numerical integration of Eq. (3.27). Poggiolini closed form solution (red line) is given by Eq. (3.40); Johannisson solution (green line) is given by Eq. (3.43); Savory closed form approximation (blue line) is given by Eq. (3.44).

Note that, for EDFA amplified Nyquist WDM systems, there exists an approximate solution [98, Eq. (23)]

$$\epsilon \approx \frac{3}{10} \log \left[1 + \frac{6}{\alpha L_{\mathrm{s}} \operatorname{asinh}\left(\frac{\pi^2 |\beta_2|}{2\alpha} B^2 \right)} \right]. \tag{3.39}$$

3.2.3 Closed-form approximations validation

In this subsection, the existing closed form approximations are thoroughly examined, and then verified via numerical Monte Carlo integration of a reference double integral Eq. (3.27). The efficiency of NL distortions due to Kerr nonlinearity is analysed as a function of the number of WDM channels (*i.e.,* transmitted bandwidth). For the dual polarisation case, lumped EDFA amplification, and (ideal) Nyquist-spaced WDM transmission system[4], the NL distortion coefficient η can be approximated using the following closed-form

[4]Nyquist criterion is assumed to be fully satisfied, *i.e.,* a width of rectangular spectra is considered to be exactly equal to the symbol rate.

solution derived by Poggiolini *et al.* [98, Eq. (41)]

$$\eta = \left(\frac{2}{3}\right)^3 \frac{\gamma^2 L_{\text{eff}}}{\pi |\beta_2| R_S^2} \operatorname{asinh}\left(\frac{\pi^2}{2} |\beta_2| L_{\text{eff}} \cdot B^2\right), \tag{3.40}$$

where L_{eff} is the effective fibre lenth defined by Eq. (2.91). Johannisson *et al.* drived the following closed form approximation [100, Eq. (16)]

$$\eta = \frac{16}{27} \frac{\gamma^2}{\pi^2 \alpha |\beta_2| R_S^2} \operatorname{Im}\left[\operatorname{Li}_2\left(\imath \left(\frac{B}{B_0}\right)^2\right)\right], \tag{3.41}$$

where $\operatorname{Li}_2(\cdot)$ is the Spence's dilogarithm, and the FWM efficiency bandwidth (a walk-off bandwidth) B_0 is given by[5]

$$B_0 = \frac{1}{\pi} \sqrt{\frac{\alpha}{|\beta_2|}}. \tag{3.42}$$

Using the asymptotic expansion of $\operatorname{Li}_2(x)$ [117], Eq. (3.41) can be approximated as

$$\eta \approx \frac{16}{27} \frac{\gamma^2}{\pi \alpha |\beta_2| R_S^2} \ln\left[\left(\frac{B}{B_0}\right)^2\right], \tag{3.43}$$

Finally, the following closed form expression was published by Savory [118, Eq. (12)]

$$\eta = \frac{16}{27} \frac{\gamma^2 \alpha L_{\text{eff}}^2}{\pi |\beta_2| R_S^2} \operatorname{Ti}_2\left(\frac{2\pi^2 |\beta_2|}{3\,\alpha} \cdot B^2\right), \tag{3.44}$$

where the introduced inverse tangent integral $\operatorname{Ti}_2(\cdot)$ defined by

$$\operatorname{Ti}_2(x) \triangleq \int_0^x \frac{\operatorname{atan}(t)}{t}\, dt. \tag{3.45}$$

Fig. 3.5 shows that for a single-channel system, an excellent agreement with numerical integration of the reference integral Eq. (3.27) has been attained by using the approximated closed form solution given by Eq. (3.44), where the SPM effect has a dominating contribution. For multi-channel Nyquist-spaced systems, both solutions Eq. (3.40) and Eq. (3.43) provide sufficiently accurate predictions. The observed small discrepancies between those closed form approximations can be clearly explained as a result of different assumptions of the integration domain. However, throughout this chapter we mainly use the numerical integration of a reference integral accounting for the NL interference,

[5] This normalisation bandwidth appears in [95] as well.

instead of using any approximated solutions given by closed form expressions. It must be noted that to keep consistency the original notations in all these closed form expressions were changed accordingly.

3.3 Digital backpropagation

One of the most popular approach to mitigate the NL interaction during signal propagation in a fibre consists in attempting to undo NL distortions after signal detection at the receiver in the digital domain by applying the inverse operators. Since, in the absence of noise, the optical pulse propagation obeys a deterministic partial differential equation, its effects can be, in principle, inverted. In other words, the transmitted electric field can be effectively reconstructed using the coherently-detected field at a generic transmission distance. This fibre nonlinearity mitigation technique is referred to as DBP and can be applied both to a single channel and to a multi-channel WDM communication system. The key ideal of DBP is to numerically solve the propagation equations (either NLSE or ME) in their inverse form, *i.e.,* with the reversed-sign fibre parameters

$$\begin{aligned} \alpha &\to -\alpha\,, \\ \beta_2 &\to -\beta_2\,, \\ \gamma &\to -\gamma\,. \end{aligned} \tag{3.46}$$

In the case when DBP is partially applied over a certain bandwidth, the contribution of deterministic S-S interactions can be evaluated as follows

$$P_{s-s} \approx N_{\mathrm{s}}^{\epsilon+1} \left[\eta(B) - \eta(B_{\mathrm{DBP}})\right] P^3, \tag{3.47}$$

where B_{DBP} stands for the DBP compensation bandwidth. Obviously, if the full signal bandwidth is applied to the DBP algorithm (*i.e.,* the so-called Full-field (FF) DBP is applied), the signal-signal NL interactions can be entirely compensated, *i.e.,* $P_{s-s} \to 0$. However, in the presence of additive noise (*e.g.,* ASE noise), complete reconstruction of the transmitted signal is not feasible due to the NL interaction between the signal and noise, as well as other stochastic effects, such as polarisation mode dispersion (see, *e.g.,* [119]). Here it should be noted that the impact of the practically relevant PMD effects on the performance of DBP is beyond the scope of this chapter. The DBP implementation accounting for the PMD has been recently considered in [120].

The next sections are devoted to analytical and numerical modelling of

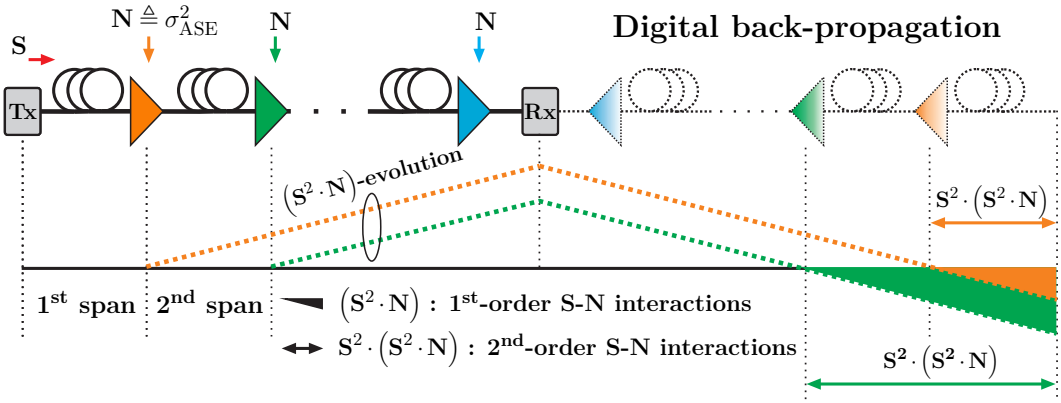

Figure 3.6: Schematic of S-N noise interactions accumulation process in an optical communication system using FF DBP, where **S** and **N** denote the signal and the ASE noise power, respectively. The triangles in a link represent lumped EDFAs, which induce linear ASE noise, whereas the inverted triangles at the receiver **Rx** side are ideal DBP amplifiers, which do not inject any noise. The dotted fibres in a virtual DBP link at a receiver side depict the fibres with negative parameters (see Eq. (3.46)). The orange and green dotted lines schematically represent the distance evolution of first-order SNI power owing to the signal interaction with noise arising from the first fibre span (with the orange EDFA amplifier) and the second fibre span (with the green EDFA amplifier), respectively. The area of filled orange and green triangles depicts the increase of uncompensated first-order SNI products (*i.e.,* $\mathbf{S}^2 \otimes \mathbf{N}$). The length of two-sided orange and green arrows delineates the growth of second-order SNI products (*i.e.,* $\mathbf{S}^2 \otimes (\mathbf{S}^2 \otimes \mathbf{N})$) due to the signal interaction with first-order uncompensated SNI noise coming from previous fibre spans in the a virtual DBP link.

nonlinearity-compensated fibre optic communication systems when DBP is either partially of fully applied.

3.4 Modelling of signal-noise interaction

Regardless of the specific scheme, when nonlinearity compensation is ideally operated over the entire transmitted bandwidth, complete suppression of all *deterministic* signal-signal interaction (SSI) can be achieved, whilst *stochastic* NL distortions consisting of signal-noise interaction (SNI) are always left uncompensated. This leads to an increased optimal SNR and optimal transmitted power compared with conventional receiver compensation schemes, such as EDC. In addition, operating beyond the optimum power may enable potential advantages, *e.g.,* in submarine transmission and for improved detection schemes [21, Sec. 10]. Within the range of powers, which are of interest for nonlinearity-compensated systems, first-order perturbation analysis is no

longer sufficient for accurate characterisation of system performance, and thus, second-order NL effects need to be taken into consideration. For instance, in [121], it was pointed out that, beyond the optimum power threshold and in the case of FF DBP, the SNR decreases with a rate of 3 SNR [dB] per power [dB], rather than 1 SNR [dB] per power [dB] as conventional first-order perturbative analytical models predict [122]. This rapid SNR degradation has been attributed to second-order SNI. This term arises from the additional NL mixing process between signal and residual first-order SNI, originating in the previous uncompensated fibre spans within a *virtual* DBP link. This is schematically illustrated in Fig. 3.6. This effect has recently been studied in [121] for single-channel optical transmission systems with a DP QPSK signal. The analysis was based on analytical closed-form expressions derived for OFDM transmission, assuming that NL interference has a Gaussian distribution as well as being fully independent of input signal modulation format. Hence, the dependence on the number of channels as well as the impact of modulation format on the effective variance of NL distortions have not been analysed. However, the implementation of a model accounting for number of channels and modulation format dependency in second-order SNI enables an accurate investigation of the system AIR, which is the most natural figure of merit in coded communication systems [61, 64, 123].

Within the framework of regular perturbation analysis [86], an analytical model for Nyquist-spaced WDM optical communication systems employing nonlinearity compensation is developed. Such a model extends the work in [121] by accounting for the modulation format dependency of second-order S-N interactions in a multi-channel transmission scenario. Different modulation formats, including DP QPSK, DP 16-QAM, and DP 64-QAM, were theoretically investigated. The SNR and optimum launched power for FF DBP schemes were analysed as a function of transmission distance, accounting for first- and second-order SNI contributions. Finally, the relevance of second-order SNIs was quantified for different transmission distances and bandwidths. The developed analytical model allows us to predict the system performance at optimum power and beyond, which gives rise to further investigations of AIRs in coded transmission systems.

The performance of a dispersion-unmanaged nonlinearity-compensated optical communication system can be evaluated by introducing the so-called *effective* receiver SNR, which includes the impact of linear ASE noise together

with NL distortions due to the optical Kerr effect, it reads

$$\begin{aligned} \mathrm{SNR} &\triangleq \frac{P}{\sigma_{\mathrm{eff}}^2}, \\ \sigma_{\mathrm{eff}}^2 &\approx N_{\mathrm{s}}\sigma_{\mathrm{ASE}}^2 + P_{s-s} + P_{s-n}, \end{aligned} \tag{3.48}$$

where P is the average optical power per channel, σ_{ASE}^2 is the overall ASE noise power at the end of each fibre span in a link, which is, in the case of the EDFA, defined by Eq. (3.11), P_{s-s} and P_{s-n} are the NL distortion powers due to SSI and SNI, respectively. The contribution of the deterministic SSI term in Eq. (3.48) is given by above-mentioned Eq. (3.26). Note that, in the case of EDC (solid lines in Figs. 3.7 and 3.8), the contribution of SNI term P_{s-n} is sufficiently small and is assumed to be negligible, whilst, for the DBP case, following the same approach as in [124], this term can be expressed as follows

$$\begin{aligned} P_{s-n} &= 3\left\{\sum_{n=2}^{N_{\mathrm{s}}}\left[(N_{\mathrm{s}}-n)^{\epsilon+1} + \left(3\tilde{\eta}P^2\sum_{m=1}^{n-1}(N_{\mathrm{s}}-m)^{\epsilon+1} + \ldots\right)\right]\right\}\sigma_{\mathrm{ASE}}^2\tilde{\eta}P^2 \\ &= 3\sigma_{\mathrm{ASE}}^2\left(\xi_1\tilde{\eta}P^2 + 3\xi_2\tilde{\eta}^2P^4 + \ldots\right) \\ &\approx 3\xi_1\tilde{\eta}\sigma_{\mathrm{ASE}}^2\cdot P^2 + 9\xi_2\tilde{\eta}^2\sigma_{\mathrm{ASE}}^2\cdot P^4, \end{aligned} \tag{3.49}$$

with ξ_1 and ξ_2 being first- and second-order distance-dependent factors, respectively, which account for the accumulation of NL optical distortions due to SNI process throughput signal propagation. Note that the pre-factor of 3 in Eq. (3.49) takes into account the assumption of equal contributions of NL scaling factors (the corresponding integrals) given by different combinations among first-order SNI products, *i.e.,* $\mathbf{S}\otimes\mathbf{S}\otimes\mathbf{N} = \mathbf{S}\otimes\mathbf{N}\otimes\mathbf{S} = \mathbf{N}\otimes\mathbf{S}\otimes\mathbf{S} = \mathbf{S}^2\otimes\mathbf{N}$, where $\mathbf{S}\triangleq P$ and $\mathbf{N}\triangleq\sigma_{\mathrm{ASE}}^2$ (see Fig. 3.6). The distance evolution of first-order SNI is captured by the factor ξ_1, which is defined as the following summation over number of fibre spans [125, Eq. (6)]

$$\xi_1 \triangleq \sum_{n=1}^{N_{\mathrm{s}}} n^{\epsilon+1}. \tag{3.50}$$

A sufficiently accurate approximated closed-form solution of Eq. (3.50), which relies on the truncation of the Faulhaber's formula [117] for $N_{\mathrm{s}} > 1$, is provided

in [125], [27, Eq. (7)]

$$\xi_1 = \frac{N_{\mathrm{s}}^{\epsilon+2}}{\epsilon+2} + \frac{N_{\mathrm{s}}^{\epsilon+1}}{2} + \frac{1}{2}\begin{pmatrix} \epsilon+1 \\ 1 \end{pmatrix} B_2 N_{\mathrm{s}}^{\epsilon} + \frac{1}{4}\begin{pmatrix} \epsilon+1 \\ 3 \end{pmatrix} B_4 N_{\mathrm{s}}^{\epsilon-2} + \dots \tag{3.51}$$

$$\approx \frac{N_{\mathrm{s}}^{\epsilon+2}}{\epsilon+2} + \frac{N_{\mathrm{s}}^{\epsilon+1}}{2}, \tag{3.52}$$

where $\begin{pmatrix} \cdot \\ \cdot \end{pmatrix}$ stands for the binomial coefficient; $B_2 = 1/6$, and $B_4 = -1/30$ are the Bernoulli numbers. The first two terms rapidly dominate the high-order contributions with increasing number of fibre spans N_{s}. Finally, the second-order factor ξ_2 is defined by the following double sum [26, Eq. (3)]

$$\xi_2 \triangleq \sum_{n=2}^{N_{\mathrm{s}}} \sum_{m=1}^{n-1} m^{\epsilon+1}. \tag{3.53}$$

Eq. (3.53) also has a closed form approximated solution. Here we provide the detailed derivations. The inner sum in Eq. (3.53) can be expressed in closed form exactly as in Eq. (3.50), and by truncating to the first two terms. Thus, we have

$$\sum_{m=1}^{n-1} m^{\epsilon+1} \approx \frac{(n-1)^{\epsilon+1}}{2} + \frac{(n-1)^{\epsilon+2}}{\epsilon+2}. \tag{3.54}$$

Therefore, we have two outer sums

$$\xi_2 = \xi_2^{(1)} + \xi_2^{(2)}, \tag{3.55}$$

where each term has the following exact analytical solution

$$\begin{aligned} \xi_2^{(1)} &\triangleq \frac{1}{2}\sum_{n=2}^{N_{\mathrm{s}}} (n-1)^{\epsilon+1} \\ &= \frac{1}{2}\left[\zeta(-\epsilon-1) - \zeta(-\epsilon-1, N_{\mathrm{s}})\right], \end{aligned} \tag{3.56}$$

and

$$\begin{aligned} \xi_2^{(2)} &\triangleq \frac{1}{2}\sum_{n=2}^{N_{\mathrm{s}}} (n-1)^{\epsilon+2} \\ &= \frac{1}{\epsilon+2}\left[\zeta(-\epsilon-2) - \zeta(-\epsilon-2, N_{\mathrm{s}})\right]. \end{aligned} \tag{3.57}$$

Table 3.1: Overview of pre-factors κ_4 and κ_6 in Eq. (3.60) for different input modulation formats and distributions

Modulation	κ_4	κ_6
QPSK constellation	1	-4
16-QAM constellation	0.68	-2.08
64-QAM constellation	0.619	-1.797
Continuous uniform distribution (∞-QAM constellation)	0.6	-1.716
Gaussian distribution	0	0

Hence, the final approximated solution reads

$$\xi_2 \approx \frac{\zeta(-\epsilon-1)-\zeta(-\epsilon-1,N_\mathrm{s})}{2}+\frac{\zeta(-\epsilon-2)-\zeta(-\epsilon-2,N_\mathrm{s})}{\epsilon+2}, \tag{3.58}$$

where $\zeta(z)$ and $\zeta(z,r)$ denote the Euler-Riemann zeta function and its extension as the Hurwitz's generalised zeta function, respectively [117]. Assuming that all the WDM channels have the same modulation format and launched power, the NL distortion coefficient in Eq. (3.49) is quantified by the following [111]

$$\tilde{\eta} = \eta - \eta', \tag{3.59}$$

where the first term η is defined by the double integral Eq. (3.27) and does not depend on the signal modulation format because the signal is assumed to be strongly dispersed, and hence, Gaussian distributed. However, the second term η' is modulation format-dependent and includes the corrections needed for non-Gaussian input signal distribution. Following the model developed in [102, 111], this modulation-dependent correction can be decomposed as follows

$$\eta' \approx \kappa_4\,\eta_1 + \kappa_4^2\,\eta_2 + \kappa_6\,\eta_3\,, \tag{3.60}$$

where the pre-factors κ_4 and κ_6 are directly related to the excess kurtosis and the sixth standardised moment of the input signal constellation, respectively. Values of κ_4 and κ_6 for typical signal distributions are provided in Tab. 3.1. Considering ideal Nyquist WDM transmission, *i.e.,* each channel has a rectangular spectrum of width exactly equal to the symbol rate, the NL scaling factors η_1, η_2, and η_3 in Eq. (3.60) have been numerically computed by means of Monte Carlo integration including both intra-channel and inter-channel effects, similar to the approach given in [107, 111]. Employing FF DBP, closed-form expressions for the optimum launched power, *i.e.,* the power corresponding to

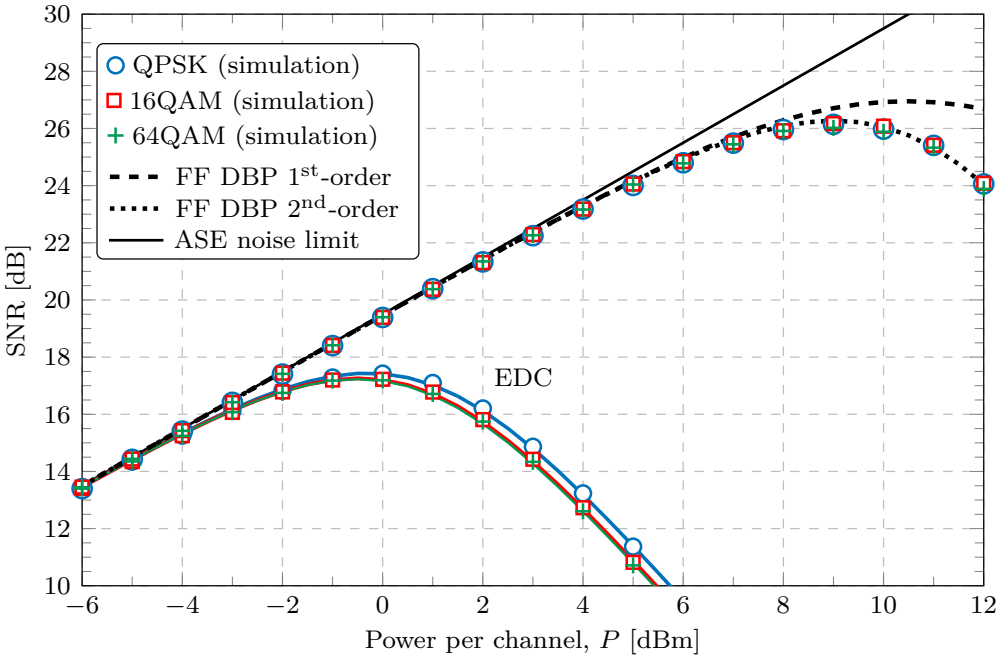

Figure 3.7: Theoretical predictions (lines) and numerical simulation results (marks) of SNR as a function of launched power per channel for a single-channel system using EDC and FF DBP. Colours refer to modulation formats.

the maximum SNR at a given number of fibre spans, for the case of first-order SNI only is given by

$$P_{\text{opt}}^{(1)} = \sqrt{\frac{N_{\text{s}}}{3\xi_1\eta}}. \tag{3.61}$$

For the second-order SNI interactions the optimum power is derived as

$$P_{\text{opt}}^{(2)} = \frac{1}{3\sqrt{2}}\sqrt{\frac{\xi_1}{\xi_2\eta}}\left(\sqrt{1+12\,\frac{N_{\text{s}}\,\xi_2}{\xi_1^2}}-1\right)^{1/2}. \tag{3.62}$$

If DBP is applied over a partial bandwidth (dashed-dotted lines in Figs. 3.7 and 3.8), the residual SNI terms P_{s-s} in Eq. (3.48) can be obtained by using an approach similar to the one in [123, 126]. It is also worth mentioning that, in the case of partial-bandwidth DBP, the NL distortions owing to residual uncompensated SSI are still more prevalent in comparison with both first- and second-order SNIs. Thus, the detrimental effect of second-order SNI can be distinguished properly in the case of FF DBP only, *i.e.*, when SSI contributions are completely suppressed.

Numerical simulations have been performed to assess the accuracy of the proposed analytical model. The investigated scenario was an ideal Nyquist-

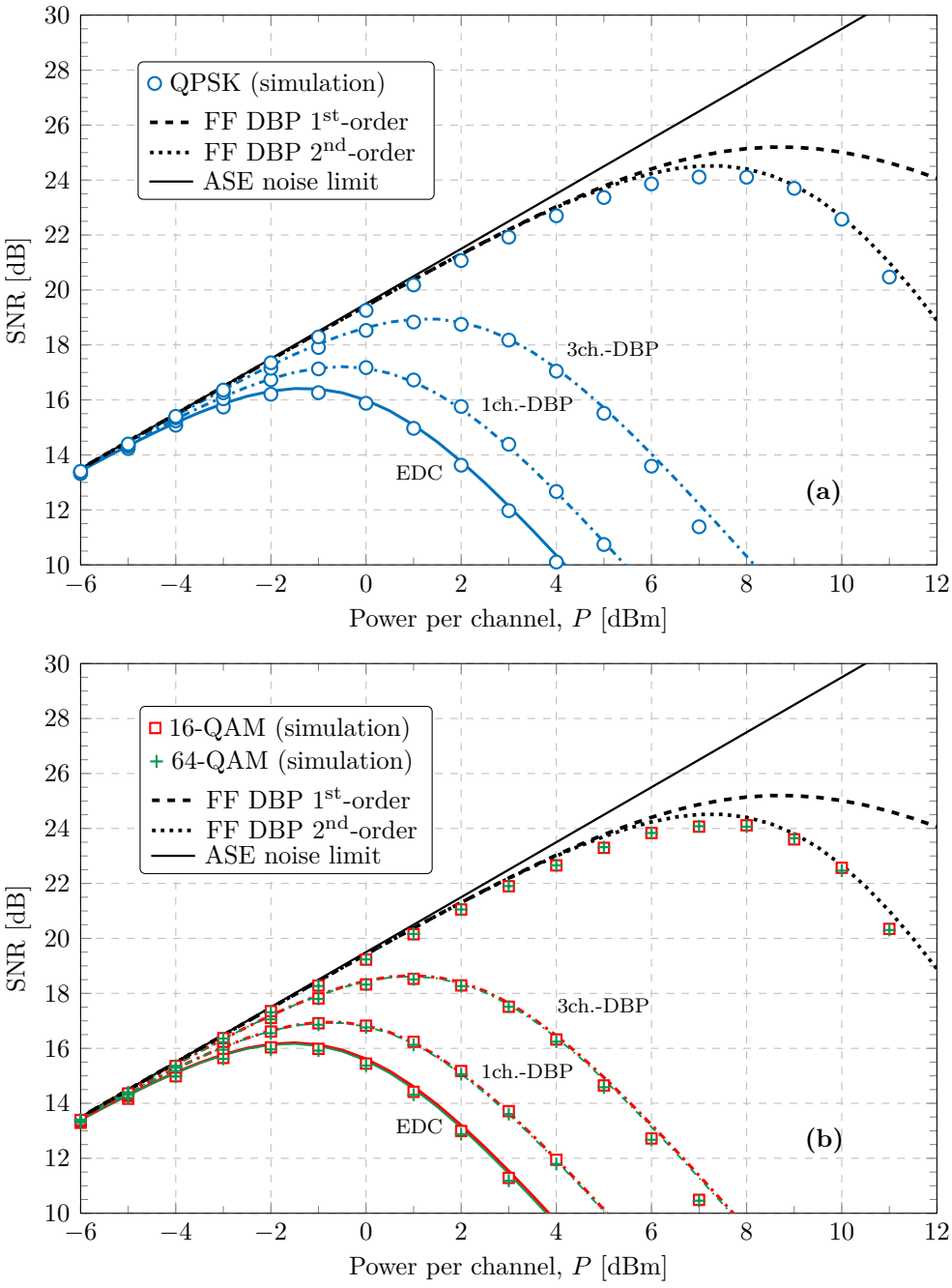

Figure 3.8: SNR performance for five-channel Nyquist-spaced WDM transmission using (a) DP QPSK and (b) DP 16-QAM, DP 64-QAM modulation formats. Colours refer to modulation formats.

spaced WDM optical fibre transmission system, using DP QPSK, DP 16-QAM, and DP 64-QAM modulation formats, with parameters shown in Table 3.2. In the transmitter, a 32 GHz spaced laser comb is employed as the phase-locked optical carrier, and the comb lines are optically demultiplexed before the I-Q modulators. A reasonable factor of 8 was used to oversample the

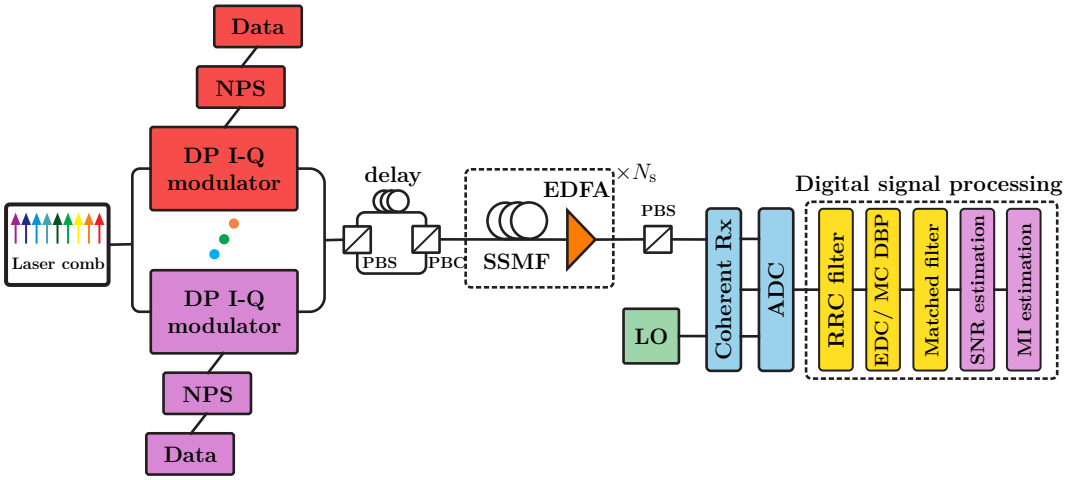

Figure 3.9: Schematic of Nyquist-spaced optical communication system using multi-channel DBP. NPS block stands for the transmitter Nyquist pulse-shaping.

Table 3.2: System parameter values

Parameter	Value
Carrier wavelength	1550 nm
Symbol rate	32 GBd
Channel spacing	32 GHz
Fibre attenuation parameter	0.2 dB km^{-1}
Fibre dispersion parameter	17 ps nm^{-1} km^{-1}
Fibre nonlinearity parameter	1.2 W^{-1} km^{-1}
Fibre span length	80 km
EDFA noise figure	4.5 dB

simulation bandwidth to obtain the best possible performance [127]. The SNR was estimated over 2^{17} symbols based on the received constellation clusters, similar to [72, 123, 128]. Note that the sequence length was chosen to ensure that the maximum "walk-off" time over the total simulation bandwidth is much shorter than the length of symbol sequence. The detailed description of simulated system set-up is illustrated in Fig. 3.9.

The transmitted symbol sequences in each optical channel are statistically independent and randomly generated, and the symbol sequences in each polarisation are identical but de-correlated with a delay of half the sequence length. The reason behind using the equal sequences on two orthogonal polarisation states rather than using independent sequences is owing to the fact that it requires less hardware memory to generate and to store the data. In other words, in order to save the resource in FPGA, the same de-correlated

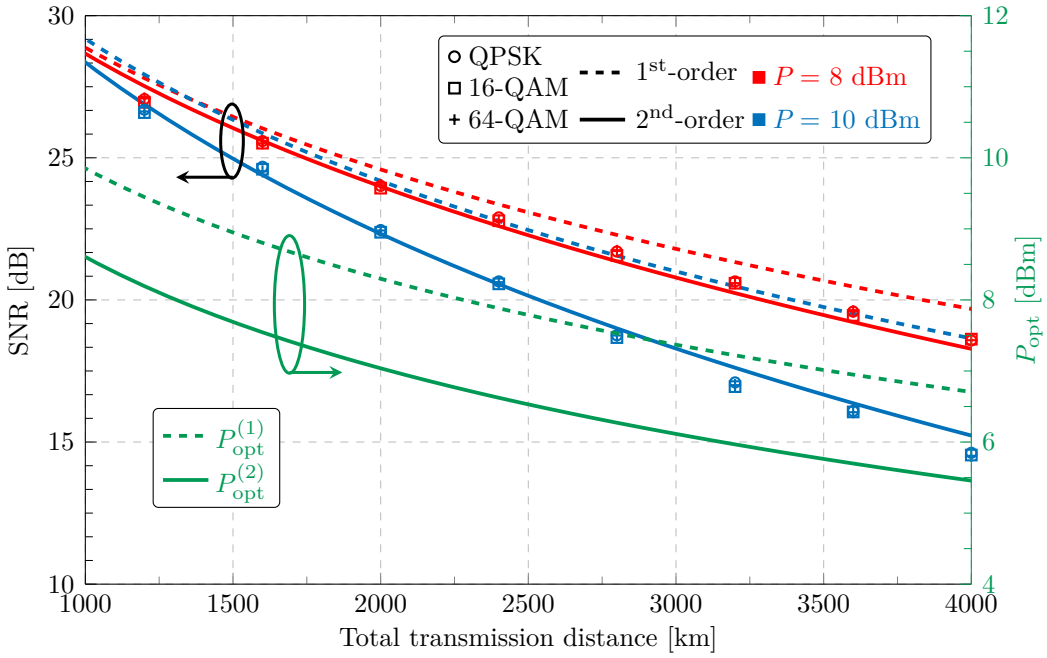

Figure 3.10: Left side: SNR versus total transmission distance at 8 dBm and 10 dBm per channel optical launched power in five-channel WDM system using FF-DBP. Numerical simulation results are indicated by marks. Right side: Distance evolution of optimum launched power per channel for FF DBP considering first- and second-order models.

symbol sequences on each orthogonal polarisation were used. The signal propagation in SSMF was simulated using the SSFT method to solve the Manakov equation, where a logarithmic step-size distribution was adopted for each fibre span, of which the mathematical expression can be found in [129, Eq. (6)]. EDFAs were employed in the loop to compensate for fibre attenuation. At the receiver, the signal was mixed with an ideal local oscillator (LO) to ensure ideal coherent detection of the optical signal. In the digital signal processing block, the EDC was implemented using an ideal frequency domain filter [68, 130], whereas multi-channel DBP was realised using the reverse split-step Fourier solution of the Manakov equation using the same step-size as in the forward propagation [66, 131] to ensure ideal operation of the multi-channel DBP. An ideal root-raised-cosine filter with a roll-off factor of 0.1 % was applied to select the desired backpropagated bandwidth for the multi-channel DBP, while no filtering was applied for the case of FF DBP. Laser phase noise and polarisation-mode dispersion were neglected.

Analytical model and numerical simulations were carried out for single- and five-channel WDM transmission systems over a typical transmission distance of 2000 km. The results are shown in Figs. 3.7 and 3.8 for the single-

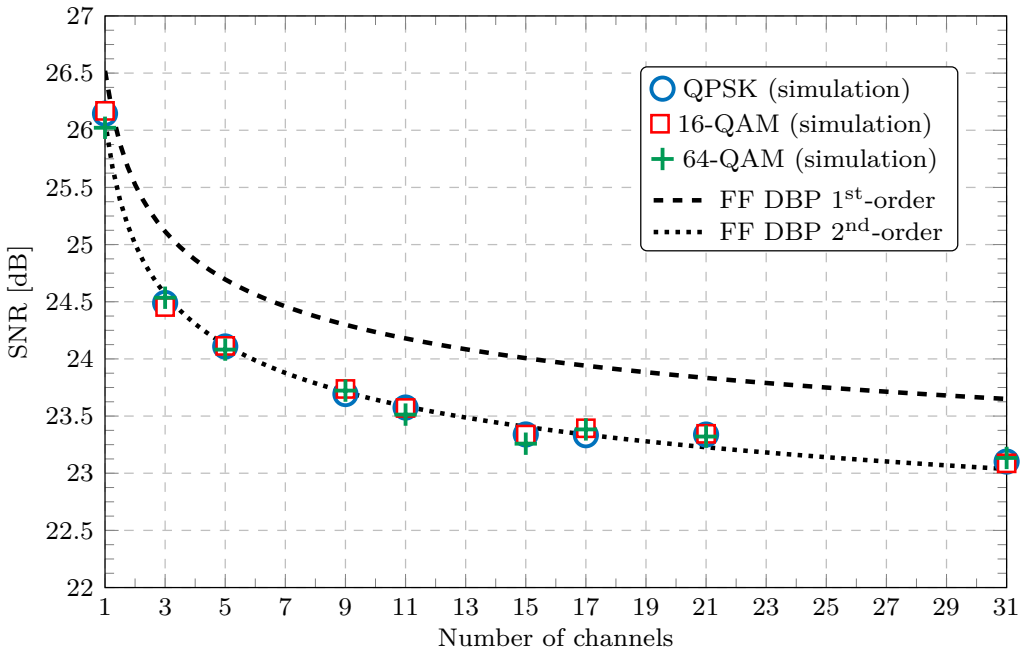

Figure 3.11: SNR against the number of Nyquist-spaced WDM channels at optimum launched power after 2000 km transmission distance using FF DBP.

channel and five-channel cases, respectively. Excellent agreement between the analytical and numerical calculations is observed. It can be seen that the second-order interactions have a significant impact on performance in the case of FF DBP, whereas it is negligible when DBP is applied over a fraction of the transmitted bandwidth. In the case of FF DBP, for values of launched power per channel beyond 7 dBm for a single-channel system (with the optimum power of 9 dBm) and beyond 5 dBm for the five-channel system (where the optimum power is 7 dBm), accounting only for first-order SNIs (dashed black lines) yields fairly inaccurate predictions. It also can be seen that the input signal modulation format does not have any substantial impact on the S-N NL distortions, when FF DBP is applied.

Fig. 3.10 indicates the variation of the SNR as a function of transmission distance and a fixed launched power in a five-channel Nyquist-spaced system using FF DBP. It can be observed that neglecting second-order SNIs leads to analytical prediction whose inaccuracy grows with both transmission distance and launched power. On the other hand, the "gap" between the predictions of optimum launched power given by the analytical models with and without accounting for the second-order interactions remains approximately constant with transmission distance. Such a gap can be quantified as approximately

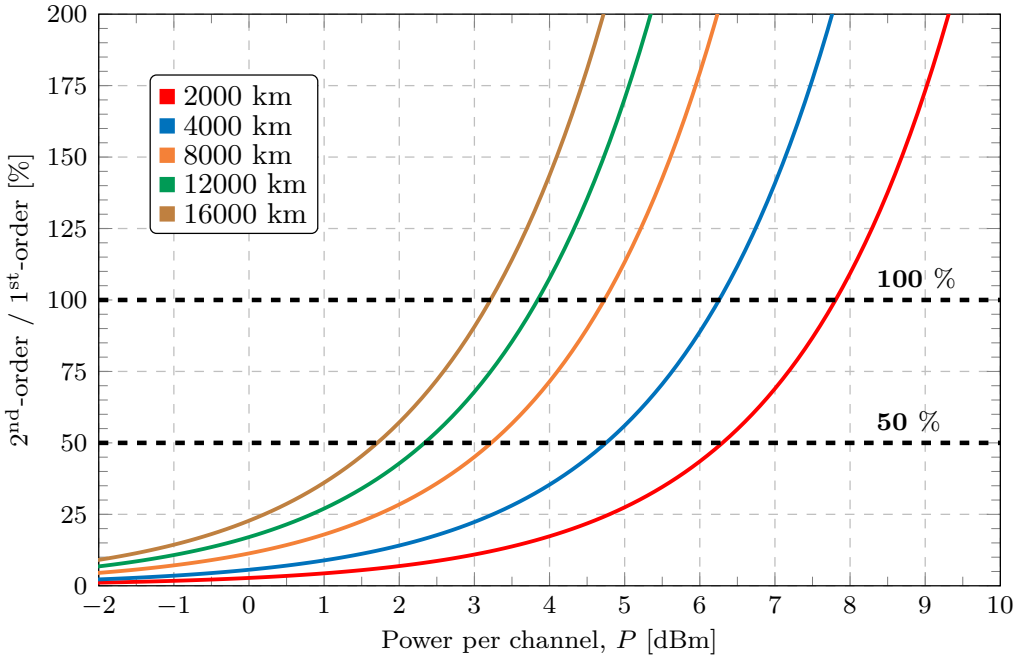

Figure 3.12: Ratio between the power of second- and first-order signal-noise FWM products against launched power per channels for different transmission distances calculated for C-band system.

1.25 dB.

This effect persists in systems with wider bandwidths. Fig. 3.11 reflects the SNR with FF DBP versus the number of transmitted channels at optimum launched power for a fixed 2000 km transmission distance. It can be seen that there is a perfect agreement between simulations and second-order SNI model was demonstrated for bandwidths of up to 31 channels (∼1 THz) and across different signal modulation formats. Because numerical simulations become computationally intractable, the analytical model is highly beneficial for predicting the SNR performance of systems with wider transmission bandwidths (*e.g.,* C-band).

Fig. 3.12 indicates the growth rate of second-order SNIs relative to the corresponding contribution of first-order SNIs as a function of launched power per channel for different transmission distances calculated for a C-band (∼4.8 THz) system using the proposed analytical model. The two horizontal dashed lines indicate the half- and equal contributions of second-order SNIs relative to first-order effects. It was found that, as both distance and launched power are increased, second-order effects become much more significant. In particular, the contribution of second-order effects grows faster with the launched power for longer transmission distances.

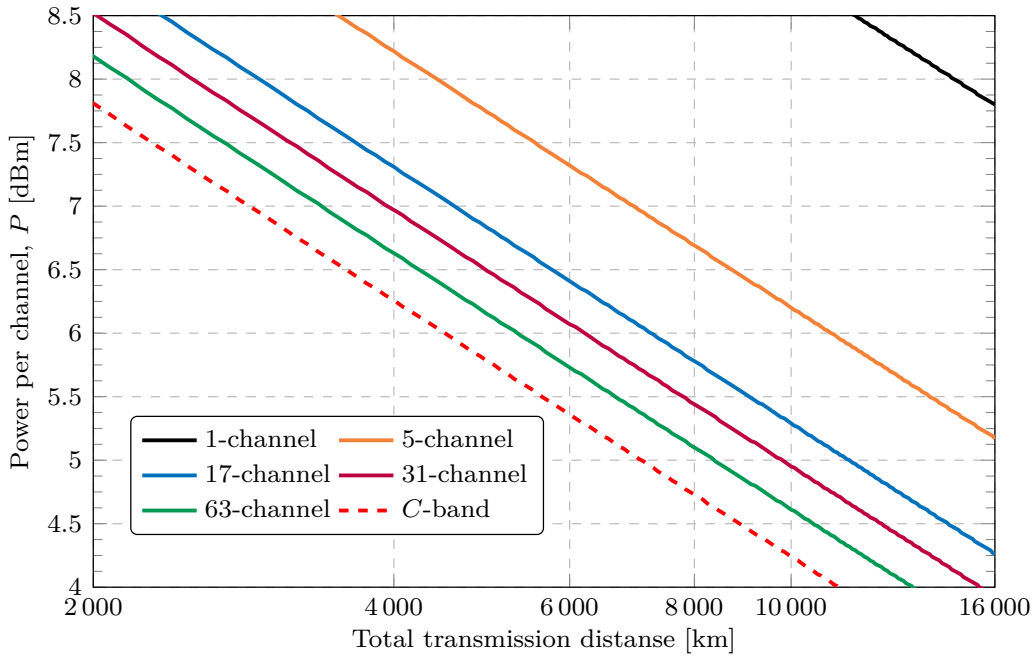

Figure 3.13: The launched power per channel corresponding to the equal contribution of the first- and second-order signal-noise interactions. The power thresholds are obtained for different bandwidths: single-channel (32 GHz), 5-channel (160 GHz), 17-channel (~ 0.5 THz), 31-channel (~ 1 THz), 63-channel (~ 2 THz) and C-band (~ 5 THz) systems.

Assuming an equal contribution of the second-order and first-order SNIs, the appropriate values of launched power per channel for different transmission distances were calculated and shown in Fig. 3.13, using different transmission bandwidths. These power values can be regarded as power thresholds beyond which second-order signal-noise interaction process in a link must be taken into account. These power thresholds can be observed to decrease as either the transmission distance or the transmission bandwidth is increased.

3.5 Summary

In summary, the impact of second-order SNIs in the presence of modulation-format-dependent NL distortions of a Nyquist-spaced WDM system with nonlinearity compensation has been theoretically studied, and accurate assessments of this effect in the NL regime beyond the optimum power were presented. Both analytical and numerical studies have been carried out for both single- and multi-channel WDM transmission systems. It was demonstrated that, in the case of FF DBP, the consideration of second-order SNIs becomes substantial both at the optimum launched power regime and beyond.

The proposed extended analytical model allowed us to accurately predict the performance of multi-channel transmission systems with complete nonlinearity compensation. Additionally, the power thresholds, corresponding to half or equal contribution of second-order signal-noise interactions relative to the first-order signal-noise effects, have been quantified in the C-band transmission system for different transmission distances, which suggest a range of launched powers when second-order SNIs become vitally important.

Despite that the aforementioned proposed model quite accurately describes the behaviour of SNR as a function of launched power beyond optimum power regime, there still exist fairly small discrepancies between analytical results and simulations in the case of FF DBP. These small deviations are mainly due to some imperfections of proposed analytical approach. In particular, the PMD effects (see, *e.g.,* [120]), as well as the influence of spectral broadening effect were not considered. The comprehensive analysis of the performance of multi-channel DBP accounting for the spectral broadening effect due to Kerr nonlinearity has been recently published in [132]. The proposed model can also be potentially extended by considering the interactions higher than second-order in a virtual DBP link.

Chapter 4

Ultra-wide bandwidth transmission modelling

To date, optical fibre communications have experienced an unprecedented growth and success over the past three decades and now stands supreme as the enabling technology that underpins the global information infrastructure. The data rates of optical communication systems have been raised from 100 Mbit/s per fibre in the 1970s to 10 Tbit/s in current commercial systems, an astonishing 100,000-fold increase. The key technologies that fuelled this surge in capacity were the above-mentioned WDM and coherent detection, improved fibre design and fabrication, as well as the possibility of optically amplifying an ultra-wideband signal [133]. The use of lumped EDFA and DRA negated the need for expensive electronic regenerators and enabled a dense WDM transmission, although the success and performance of these amplifier technologies are now viewed as limiting the usable fibre bandwidth to approximately 10-15 THz, ultimately limiting the maximum throughput of optical network systems.

As explained in Sec. 1.2, attaining higher information rates is essential to cope with the ever-growing data demand, directly related to the well-known "capacity crunch" issues (see Sec. 1.2). In coded transmission systems, AIR is an important figure of merit as it corresponds to the net data rate that can be achieved by a transceiver based on the so-called *soft-decision* decoding [61, Sec. IV], [134]. Clearly, for a fixed signal bandwidth, higher AIRs (or throughputs) are commonly achieved by using closer channel spacing (*e.g.,* Nyquist-spaced WDM systems) and denser signal constellations. However, higher cardinality of signal modulation formats typically operate well only at high SNRs, that is, where high optical launched power is required [135]. In such scenarios, the AIR of optical fibre communication systems is inherently limited by the aforementioned NL distortions owing to the optical Kerr effect in a fibre, which,

in the frequency domain, can manifest itself as SPM, XPM and FWM. Much research has been devoted to improving these limits through a variety of NL compensation techniques, such as DBP, optical phase conjugation, twin-wave phase conjugation, NFT, and others, as recently reviewed in [29]. Also, AIRs can be increased by means of signal shaping [70, 136, 72]. Geometric shaping enhances the AIR by changing the uniform signal constellation grid, whereas probabilistic shaping (PS) utilises a non-uniform probability distribution on the constellation points. In this chapter, we focus on probabilistic shaping since it offers several advantages over geometric shaping, as discussed in [72, Sec. I].

4.1 Ultra-wide bandwidth AIR estimations

Investigations of the AIRs for C-band and beyond rely on analytical calculations since the computational complexity of split-step simulations makes them infeasible. The AIRs have already been evaluated for an EDFA system with a total bandwidth of 5 THz [98] considering EDC case only. Additionally, AIRs have been investigated for a bandwidth of 4.3 THz [87], assuming IDRA and considering EDC only. In all the reported works, the NL distortion coefficient of the central channel has been employed as the NL distortion coefficients for all channels over the entire optical bandwidth to estimate the AIRs [87, 137, 98]. However, these coefficients are smaller in outer channels due to lower inter-channel nonlinearities.

As mentioned above, all conventional models are mainly under the assumption of the flatness of noise-like NL distortions spectrum, *i.e.,* every frequency component experiences exactly the same power evolution along the link. This is a reasonable assumption unless an ultra-wide bandwidth transmission system is under consideration. Avoiding the assumption of the flatness of the NLI spectrum, the SNR at the receiver of each k individual channel can be defined as[1]

$$\begin{aligned} \mathrm{SNR}\,(k) \triangleq \frac{P}{\sigma^2_{\mathrm{eff}}\,(k)} &\approx \frac{P}{N_{\mathrm{s}}\,\sigma^2_{\mathrm{ASE}} + P_{s-n}\,(k) + P_{s-s}\,(k)}\,, \\ P_{s-s}\,(k) &= N_{\mathrm{s}}^{1+\epsilon}\,\eta(k)\,P^3\,, \\ P_{s-n}\,(k) &\approx \xi_1 \eta(k)\,\sigma^2_{\mathrm{ASE}}\,P^2\,, \end{aligned} \tag{4.1}$$

where the factor ξ_1 is responsible for the distance evolution of first-order SNI and defined by Eq. (3.50), and the NL distortion coefficient η of WDM Nyquist-

[1] Here we omit the second-order SNI contributions.

spaced k-channel is

$$\eta(k) = \frac{1}{R_S} \int_{(\frac{2k-1}{2})R_S}^{(\frac{2k+1}{2})R_S} d\nu \, S(\nu) \, H(\nu) \, . \tag{4.2}$$

where the Nyquist WDM channel index k is given by the following set $k \triangleq \{-(N_{\text{ch}}-1)/2, \ldots, (N_{\text{ch}}-1)/2\}$. Eq. (4.2) implies the filtering of the PSD $S(\nu)$ at the coherent receiver by a matched filter with a baseband transfer function $H(\nu)$. The PSD $S(\nu)$ of NL noise-like distortions is given by

$$S(\nu) = \frac{16\gamma^2}{27R_S^2} \int_{-\frac{B}{2}}^{\frac{B}{2}} \int_{-\frac{B}{2}}^{\frac{B}{2}} d\nu_1 d\nu_2 \, |\mu(\nu_1, \nu_2, \nu)|^2 \, \text{rect}\left(\frac{\nu_1 + \nu_2 - \nu}{B}\right), \tag{4.3}$$

where the appropriate FWM efficiency factor $\mu(\nu_1, \nu_2, \nu)$ is defined precisely as in Eq. (3.28), with the FWM phase-mismatch $\Delta\beta$ being

$$\begin{aligned} \Delta\beta(\nu_1, \nu_2, \nu) &= \beta(\nu_1 + \nu_2 - \nu) - \beta(\nu_1) - \beta(\nu_2) + \beta(\nu) \\ &\approx 4\pi^2 \, |\beta_2| \, (\nu_1 - \nu)(\nu_2 - \nu) \, . \end{aligned} \tag{4.4}$$

Typical fibre values (SSMF) were used to evaluate the system performance with the parameters shown in Tab. 4.1. The ASE noise at the receiver in Nyquist WDM systems was calculated according to Eq. (3.11) for EDFA and Eq. (3.19) for DRA.

Given its utmost precision in long-haul, highly-dispersive fibre-optic communication systems with dense signal modulation formats, a first-order perturbation analysis was used to compute the NL distortion coefficients, *i.e.,* the measure of noise-like NL interference [87, 98, 99, 61, 137]. The assumption of Gaussian signal distributions overestimates the impact of NL distortions with regard to a uniform QAM signal constellation. In fact, this holds for most probabilistically-shaped QAM constellations and, thus, a Gaussian distribution provides lower bounds on the SNR for most PMFs, including the Maxwell-Boltzmann discrete input distribution. Therefore, all calculated AIRs are guaranteed to be achievable.

The impact of dispersion slope (β_3) is neglected in this section. The β_3 will be implemented in the model in the next section, as well as its influence on the NL distortion spectrum and on the achievable rate will be analysed. The whole spectrum of the NL distortion coefficients is calculated for both

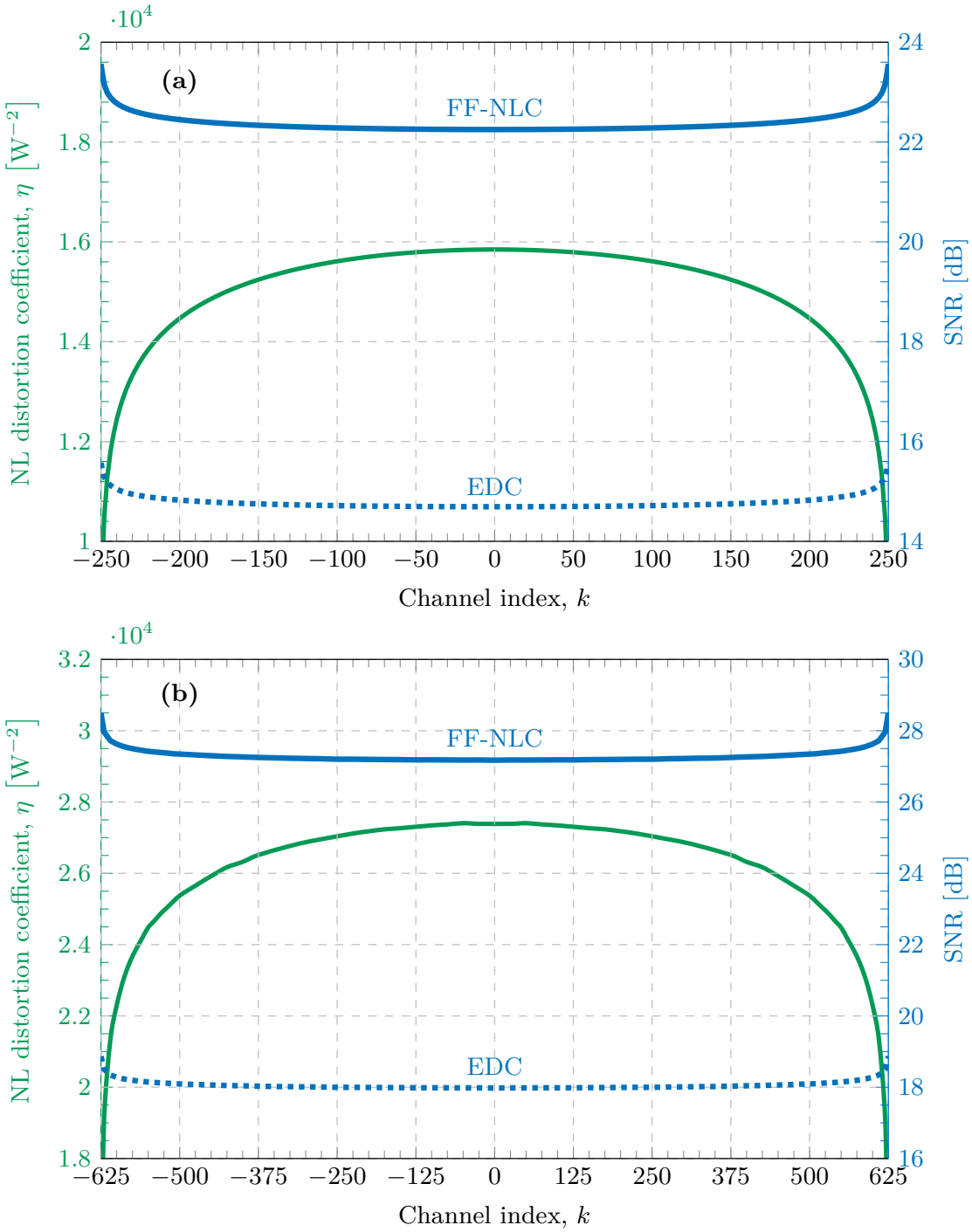

Figure 4.1: NL distortion coefficients and SNR values at 2000 km 25×80 km. (a): EDFA system and (b): Raman-amplified system.

discussed amplification schemes and is shown in Fig. 4.1.

In addition, the SNR according to Eq. (4.1) is plotted for EDC only (*i.e.*, $P_{s-n} = 0$) and for FF-NLC (*i.e.*, $P_{s-s} = 0$) at optimum and uniform launched power across the spectrum. For EDC only, the SNR at optimum

Table 4.1: System parameter values

Parameters	Values
Carrier wavelength	1550 nm
Symbol rate	10 Gbaud
Channel spacing	10 GHz
Roll-off factor	0%
EDFA, Num. of channels,	501
DRA, Num. of channels,	1251
Total Raman pump power	5×680 mW
EDFA NF	4.5 dB
Fibre span length	80 km
Fibre loss	0.20 dB km^{-1}
Fibre loss for Raman pump beam	0.25 dB km^{-1}
Fibre dispersion	17 ps nm^{-1} km^{-1}
Fibre dispersion slope	0 ps nm^{-2} km^{-1}
Fibre NL coefficient	1.2 W^{-1} km^{-1}
Line-width of transmitter	0 Hz
Line-width of local oscillator	0 Hz

input power scales as follows

$$\Delta\mathrm{SNR}_{\mathrm{NLC}}\,[\mathrm{dB}] \approx -\frac{1}{3}\,\eta\,[\mathrm{dB}] \tag{4.5}$$

whereas, for FF-NLC, it scales as

$$\Delta\mathrm{SNR}_{\mathrm{NLC}}\,[\mathrm{dB}] \approx -\frac{1}{2}\,\eta\,[\mathrm{dB}] \tag{4.6}$$

Therefore, the change in SNR is slightly different at the edges of the spectrum in the FF NLC case, compared to the EDC-only case. For the central channel, Raman amplification outperforms EDFA amplification by 3.2 dB and 4.9 dB for EDC only and FF-NLC, respectively. The bigger margin using FF-NLC is easily understood by looking at the scaling of the FF-NLC gain which is approximately given by

$$\Delta\mathrm{SNR}_{\mathrm{gain}}\,[\mathrm{dB}] \approx -\frac{1}{6}\,\eta\,[\mathrm{dB}] - \frac{1}{3}\left(N_{\mathrm{s}}\,\sigma^2_{\mathrm{ASE}}\right)[\mathrm{dB}]\,. \tag{4.7}$$

This relation yields a ΔSNR of 1.63 dB in favour of Raman amplification since its ASE noise contribution is lower (see Fig. 3.3).

In order to compute the AIR, the soft-decision MI for each channel was numerically calculated using the Gaussian-Hermit quadrature (see Eq. (2.125)). Here, the MI is a transmission rate that can be achieved, assuming a Gaussian

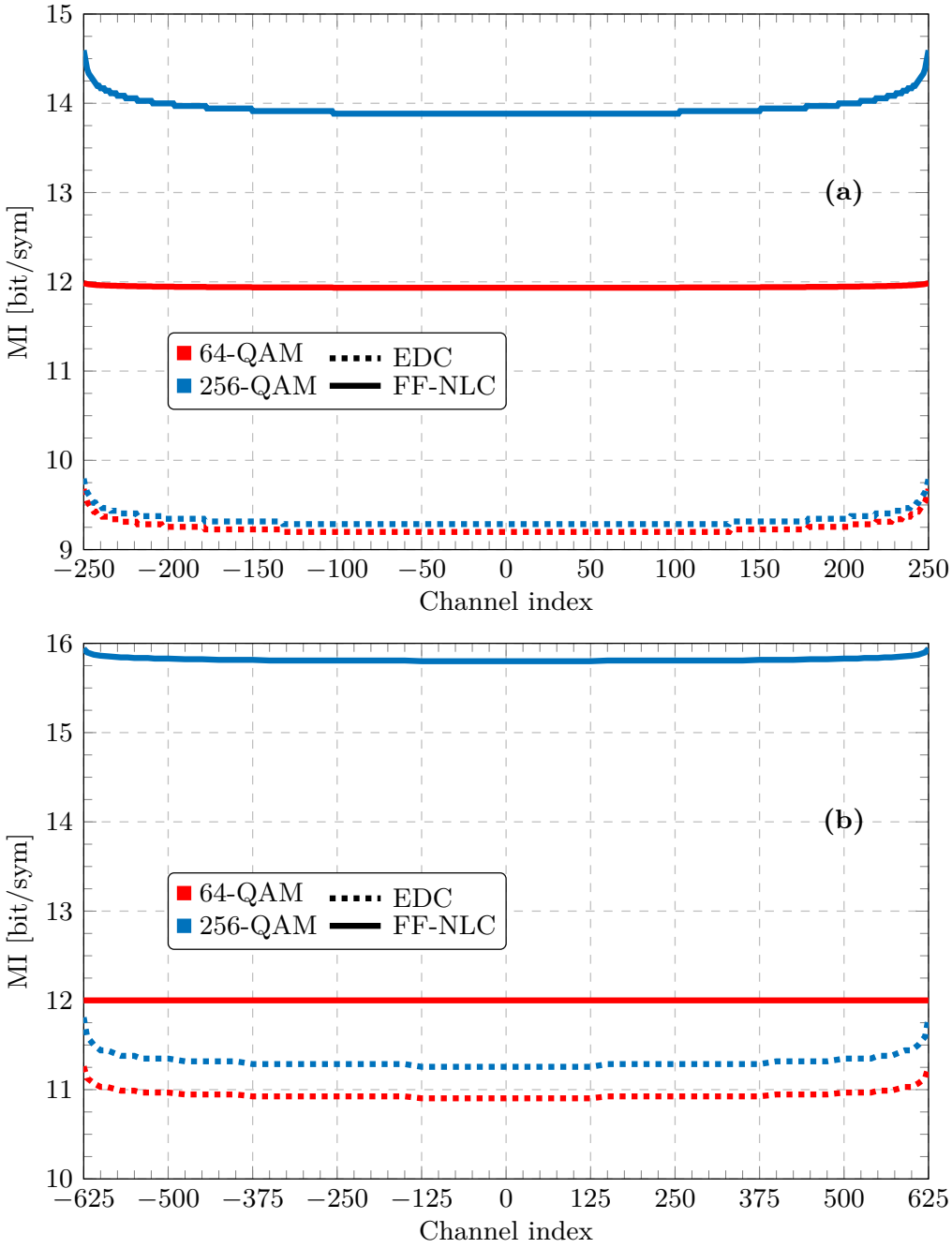

Figure 4.2: Mutual information for each channel at 2000 km 25×80 km. (a) EDFA: system and (b): Raman-amplified system.

channel law, *i.e.*, leading to the following approximations for the effective noise variance in Eq. (2.112), that is, $\sigma^2_{\text{eff}} \approx \sigma^2_z$. Fig. 4.2 shows the results for both amplification schemes.

For the transmission length of 2000 km and the EDC-only case, the use of DP 64-QAM is a good compromise between the performance and complexity as

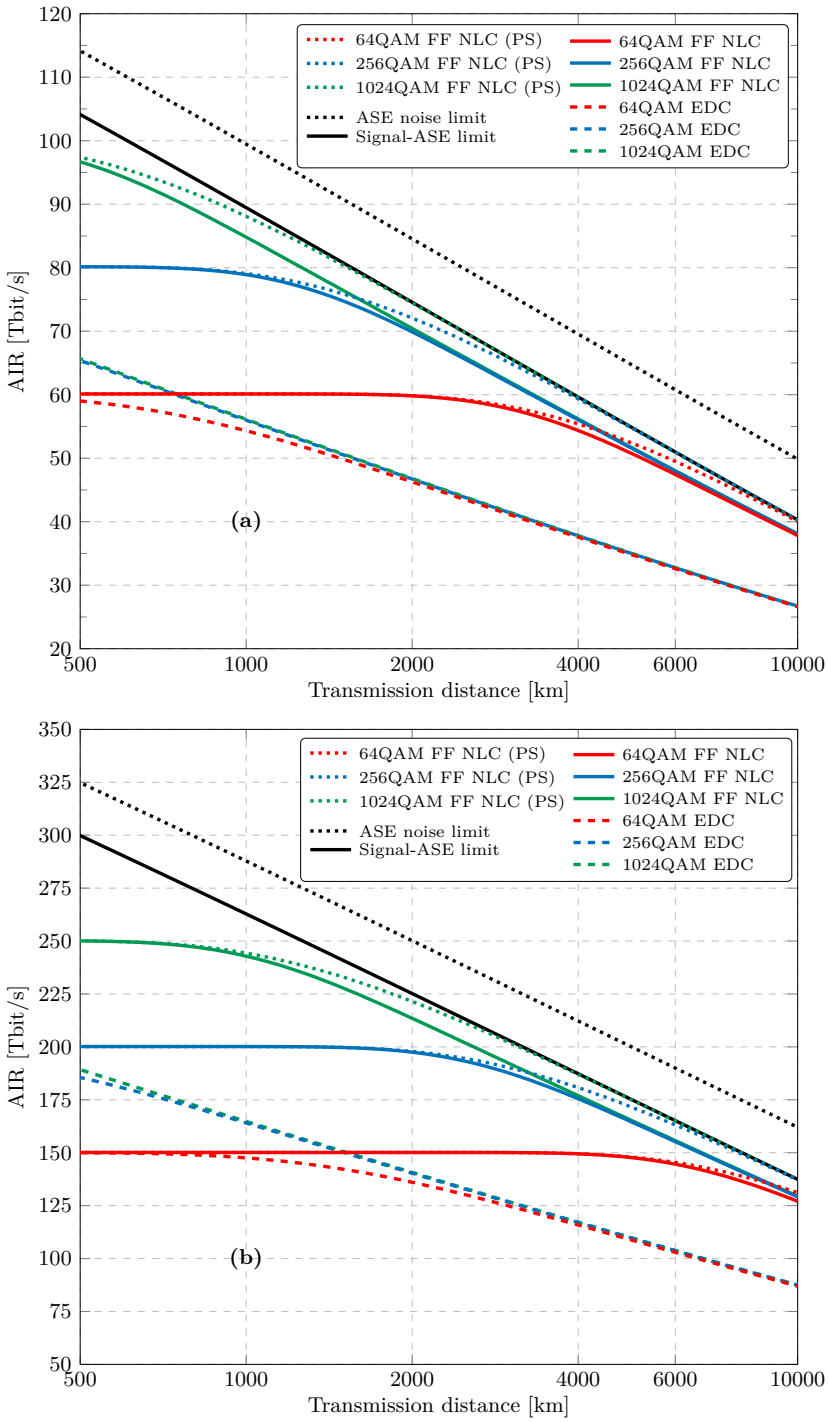

Figure 4.3: AIRs of optical transmission systems. (a): EDFA system and (b): Raman-amplified system. PS stands for the probabilistically-shaped signal constellation.

DP 256-QAM gives marginal improvement. However, when FF NLC is applied, the MI for DP 64-QAM saturated at 12 bit per symbol. Higher modulation formats need to be applied such as DP 256-QAM, as it would yield a MI of

∼14 bit per symbol for the EDFA scheme and ∼16 bit per symbol for DRA scheme.

To further increase the MI, a probabilistically-shaped constellation was applied. The idea is to transmit constellation points with a non-uniform probability according to a certain PMF. More information on PS and typical constellation diagrams can be found in [72, 70]. For a Gaussian noise channel model under an average power constraint, it can be shown that signal shaping yields a gain of up to 1.53 dB, which is referred to as the well-known ultimate shaping gain (see, Sec. 2.3.2). As the optimum input distribution for the NL fibre-optic channel is still under debate, a Maxwell-Boltzmann distribution was suggested, since it has been proved to be optimal for QAM constellations in a Gaussian channel [136]. This is fully consistent with the first-order perturbation analysis of NL distortions.

An important property of NL channels is the SNR dependence on the input signal distribution. However, as mentioned above, the SNR calculated based on a Gaussian distribution can be represented as a lower bound. In order to obtain a tighter lower bound, modulation format dependent models ought to be considered (see, *e.g.,* [107, 138]), as was done in the previous section. In the EDFA case, an approximated closed-form expression is available that approximately corrects for the modulation dependence of cross-channel interference [138], however, this solution is not applicable to examine the performance of each WDM channel individually.[2] Applying the used fibre-optic system parameters, the expression yields a higher optimal SNR of 0.63 dB for DP-64QAM. For higher modulation formats, this difference will be slightly smaller. These small corrections (strictly positive) are neglected as the modulation format dependent model will introduce unmanageable additional complexities.

In the following, both EDFA and DRA schemes are presented in terms of AIRs for three cases, namely EDC-only, FF-NLC, and PS combined with FF-NLC. The results are shown in Fig. 4.3. A signal-ASE noise interaction limit is shown by employing the Shannon formula Eq. (2.113)

$$C_{s-n} = 2R_S \sum_k \log_2 \left(1 + \mathrm{SNR}_{\mathrm{FF-NLC}}(k)\right), \tag{4.8}$$

assuming a Gaussian channel law as well as Gaussian statistics of the noise due to signal-ASE interference. Similarly, a limit only considering the ASE

[2] An exact correction due to modulation format dependence can be computed via numerical integration, as it is done in the previous section.

noise is also plotted out by computing as follows

$$C_{\mathrm{ASE}} = 2R_S \sum_k \log_2 \left(1 + \mathrm{SNR}_{\mathrm{ASE}}(k)\right). \tag{4.9}$$

The use of FF NLC significantly increases the AIRs, justifying the current research efforts into NL compensation techniques. Additionally, the AIRs of the Maxwell-Boltzmann shaped input distribution approach the signal-ASE noise interaction limit for longer distances. For both EDFA and Raman-amplified fibre-optic systems, the increase in modulation formats gives a considerable improvement in the AIRs, when the transmission distance is less than 2000 km. For DP 1024-QAM and a total transmission distance of 2000 km, the application of FF NLC can realise AIRs of ∼70 and ∼215 Tbit/s for the EDFA and the Raman amplification schemes, respectively. This can be pushed further to the signal-ASE noise interaction limit by using PS that yields AIRs of ∼75 and ∼223 Tbit/s, respectively.

In addition, it was found that DP 256-QAM can achieve the same AIRs as DP 1024-QAM (both in the signal-ASE noise limit) with the use of FF NLC and PS, when the transmission distance exceeds 3200 km in the EDFA system and 6000 km in the Raman-amplified system. This implies that DP 256-QAM is sufficient over these long-haul transmission distances.

It should be noted that this model has been applied to a recent experimental record reporting 49.3 Tbit/s transmission over 9100 km using C-band L-band EDFA [139]. According to our estimation, and neglecting the ∼4 nm gap between the C-band and L-band, this record reaches ∼76% of the theoretical AIR for DP 16-QAM and EDC only, due to practical limitations. However, overcoming these practical limitations, the use of DP 256-QAM, FF NLC, and PS, would potentially double the reported transmission rate and achieve the signal-ASE interaction limit within the same bandwidth.

It is also worth mentioning that the phase noise contributions of both transmitter and local oscillator were neglected. In practical transmission systems, the phase noise will interact with dispersion compensation modules in both linear and NL compensation schemes, which may give rise to the so-called equalisation-enhanced phase noise [140, 141]. The additional impact from this effect can be the subject of future work.

4.2 Impact of dispersion slope

In this section, the AIR of optical communication systems using ultra-wideband EDFA (∼ 5 THz) as well as DRA (∼ 12.5 THz) for a SSMF trans-

mission has been numerically evaluated considering the impact of third-order dispersion (dispersion slope). Because of the expected difference in performance, due to the impact of dispersion slope, the estimations have been carried out separately for each sub-channel in Nyquist-spaced WDM transmission system.

The MI Eq. (2.107) is said to be a *maximum* information rate of a coded modulation scheme, which essentially encodes a stream of N_b information bits, *i.e.,* $\boldsymbol{b} = \left[b_1, b_2, \ldots, b_{N_b}\right]$ into a sequence of N_s symbols $\boldsymbol{x} = [x_1, x_2, \ldots, b_{N_s}]$, each drawn from a set of complex values $\mathcal{X} = [s_1, s_2, \ldots, s_{2^m}]$, where $m = \log_2|\mathcal{X}|$ is the number of bits per symbol for the signal constellation $\mathcal{X}$ with cardinality $|\mathcal{X}|$. Therefore, assuming *ideal* coded modulation encoder (*i.e.,* no loss of entropy), the coded rate $R^* \in [0,1]$ and overhead $\mathrm{OH}^* \in [0,\infty)$ are given by [134, Eq. (1)]

$$R \triangleq \frac{N_b}{N_s} \leq R^* = \frac{1}{m} I_{X,Y}\,, \tag{4.10}$$

and

$$\mathrm{OH} \triangleq \frac{N_s - N_b}{N_b} \geq \mathrm{OH}^* = \frac{1}{R^*} - 1\,. \tag{4.11}$$

In order to estimate the NL distortion power and eventually SNR at the receiver side, in the presence of the dispersion slope, the FWM phase-mismatch in the integral Eq. (4.3) needs to be modified as follows

$$\begin{aligned} \Delta\beta &= \beta(f_1 + f_2 - f) - \beta(f_1) - \beta(f_2) + \beta(f) \\ &\approx 4\pi^2 (f_1 - f)(f_2 - f)\left[|\beta_2| + \pi (f_1 + f_2)\beta_3\right], \end{aligned} \tag{4.12}$$

where $\beta(\cdot)$ denotes the propagation constant as a function of frequency, and β_2 and β_3 represent the second- and third-order dispersion coefficients in the Taylor expansion, respectively. In our investigation, we examined the backward-pumped geometry of the DRA scheme neglecting the Raman pump depletion effect as well as assuming the independence of the Raman gain on the laser wavelength. The optical lumped EDFA-amplified scheme was also analysed. Accounting for the third-order dispersion, the NL distortion coefficient η is estimated for each WDM channel individually based on the first-order perturbation solution. Note that, the rate of convergence of the reference triple integral decreases with increasing the transmitted bandwidth. Therefore, the triple integral in Eq. (4.2) was evaluated numerically via quasi-Monte Carlo integration [110]. The SNR has been evaluated at a fixed 25×80 km transmission distance. Fig. 4.4 shows the results using single-channel DBP (32 GHz)

Table 4.2: System parameter values

Parameters	Values
Carrier wavelength	1550 nm
Symbol rate	32 GBd
Channel spacing	32 GHz
EDFA bandwidth	40 nm
EDFA noise figure	4.5 dB
DRA bandwidth	100 nm
Total Raman pump power	5×680 mW
Attenuation	0.2 dB km^{-1}
Dispersion	17 ps nm^{-1} km^{-1}
β_2	– 21.67 ps^2 km^{-1}
Dispersion slope	0.067 ps nm^{-2} km^{-1}
$\beta 3$	0.145 ps^3 km^{-1}
Nonlinearity	1.2 W^{-1} km^{-1}
Transmission distance	25×80 km

and a practically feasible multi-channel DBP with 250 GHz back-propagated bandwidth [142]. The impact of β_3 is reflected in the tilt in NL distortion spectrum, because of low frequency components exhibiting a higher amount of dispersion, which results in a lower influence of NL interference. At the same time, with increasing launched powers, low frequency components are amplified at the cost of high frequency components, leading to enhanced NLI process for low frequency range. This eventually gives rise to a different performance of each channel in the WDM transmission system.

Assuming a Gaussian channel law, the code rates Eq. (4.10) (and the overhead Eq. (4.11)) considering two different signal modulation formats, such as DP 64-QAM and DP 256-QAM, were calculated for each individual WDM channel by means of the Gauss-Hermite quadrature method (see Sec. 2.3.2, more details about a numerical method can be found in [143, 144]). The detailed parameters of the optical communication system are listed in Tab. 4.2. The inclusion of the dispersion slope breaks the symmetry in the NLI spectrum, and thus, raises a question of separate optimisations for the different channels in dense WDM transmission systems. The need to optimise the code rate (or the overhead) of the coding scheme for each individual in WDM transmission system has been clearly shown in Fig. 4.5, and Fig. 4.6 throughout the section.

The most straightforward approach to optimise the performance across the whole WDM bandwidth in ultra-wideband transmission systems can be realised in terms of using adaptive signal modulation formats with respect to

each WDM channel. More involved approach can be based on implementing the so-called power adaptation algorithms, for instance, a sort of iterative water-filling algorithms (see, *e.g.,* [145, 146]) that potentially optimise the allocation of power to various WDM channels.

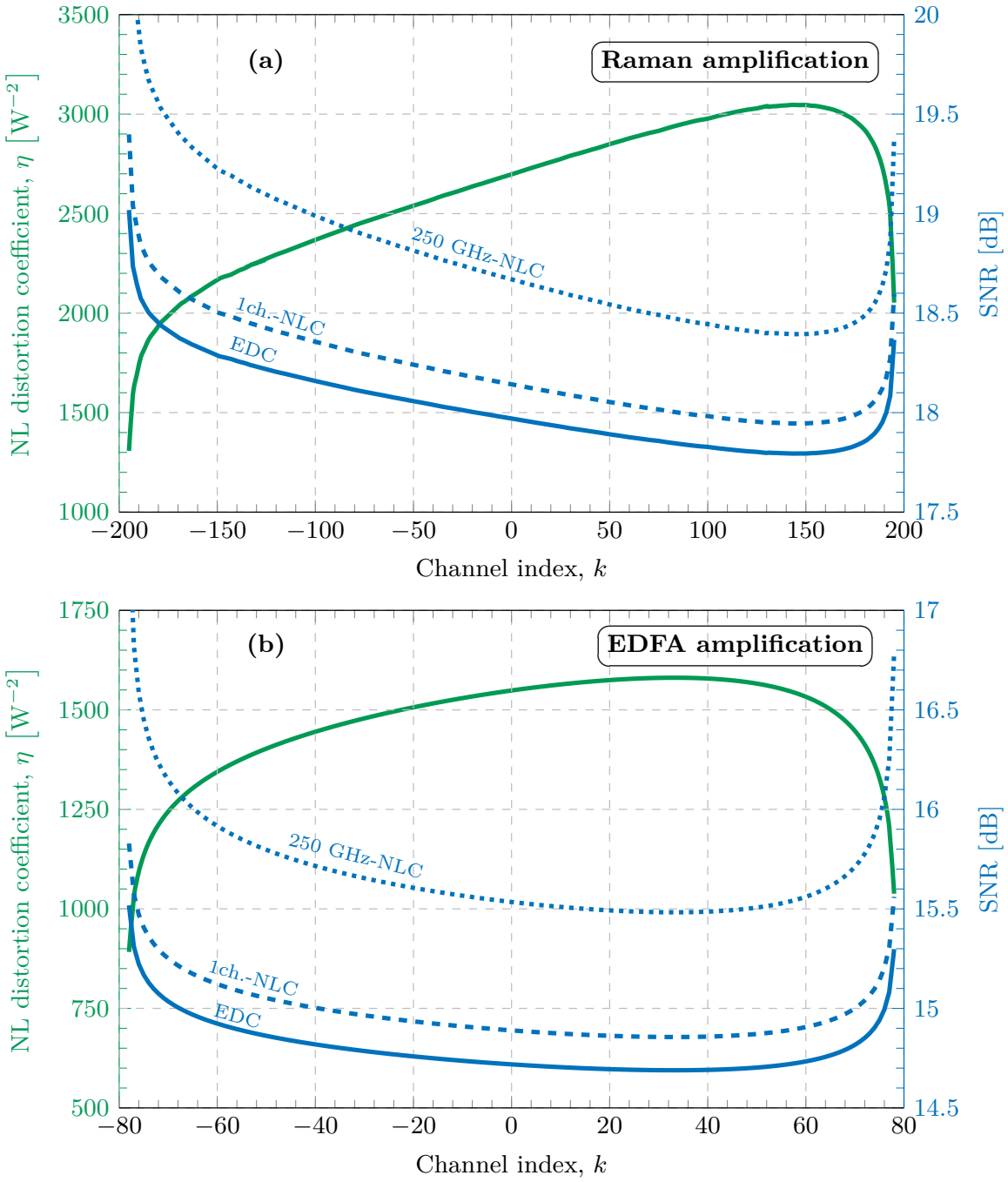

Figure 4.4: NL distortion coefficient η and SNR at fixed 2000 km (25×80 km) transmission distance against WDM channel index. (a) distributed Raman-amplified $C+L$ -band (∼12.5 THz) system and (b): lumped EDFA-amplified C-band (∼5 THz) system.

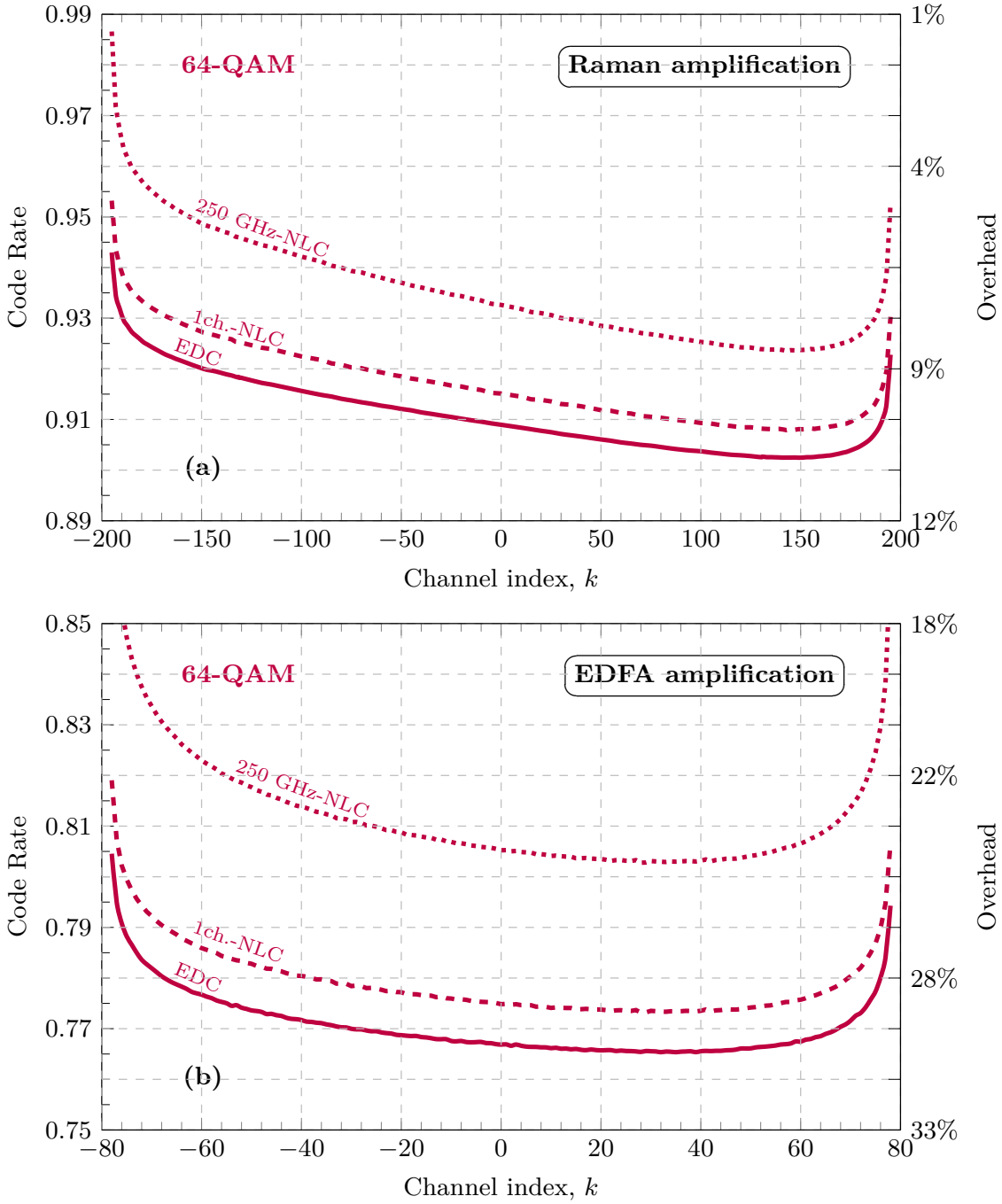

Figure 4.5: The code rate (overhead) as a function of WDM channel index for 64-QAM modulation format.

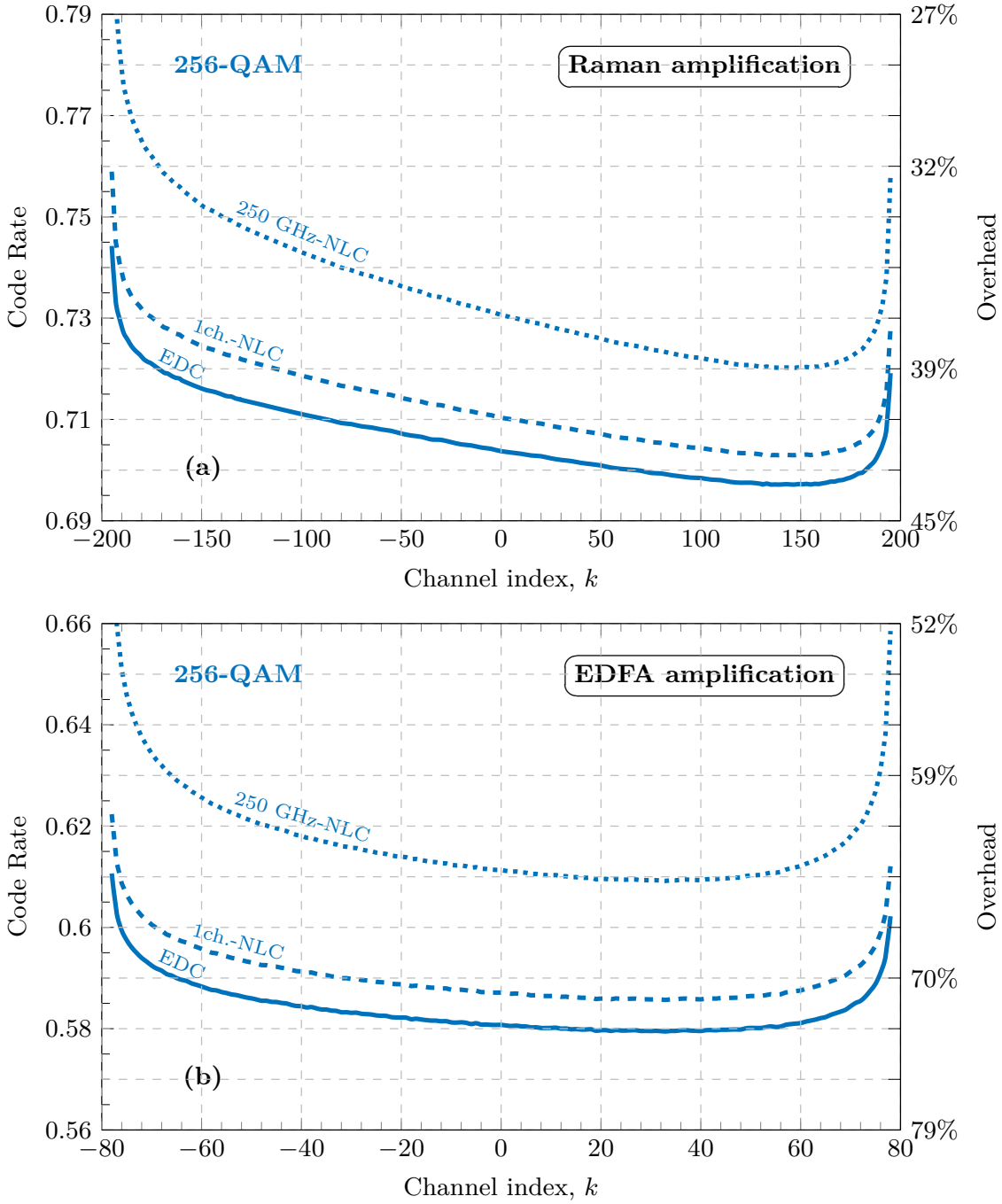

Figure 4.6: The code rate (overhead) as a function of WDM channel index for 256-QAM modulation format.

4.3 Future capacity trends

It is now beyond any doubt that there are a number of methods, which are moderately effective in mitigating the NL impairments due to the Kerr effects in a fibre [29]. The main benefit from these nonlinearity-compensation techniques appears to be in increasing the transmission distance, where the AIR can be maintained over as much as double distances in comparison with a linearly-compensated systems. Some astonishing experimental results for nonlinearity-compensated fibre-optic communication systems are shown in [131]. Nonetheless, the issues regarding the overall possible increases in the maximum fibre AIR as well as the issues regarding the future capacity trends of fibre-optic communication systems still remain open research challenges.

Considering an unrepeated link (*i.e.,* the link which does not contain any optical amplifiers), the capacity can be defined by the quantum noise limit (*i.e.,* the Poisson shot noise of the detector [147]). It has been recently asserted that for any amplified link, an upper bound on the AIR Eq. (2.135) is determined by the Shannon expression Eq. (2.113) with an equivalent AWGN channel Eq. (2.110), where the SNR is merely confined by the ASE noise arising due to optical amplification [148, 149]. This result implies that the capacity of the optical fibre channel is never above Shannon's $\log(1+\mathrm{SNR})$ expression, and that the dispersion and nonlinearity compensation do not improve channel capacity. The straightforward implication of this result means that the best we can ever hope for is to fully compensate the deterministic contribution of Kerr nonlinearity (*i.e.,* the SSI due to the SPM) and the chromatic dispersion, resulting in a linear AWGN channel, and under this assumption the nature of logarithmic function declares that increasing launched power in a fibre will always lead to progressively smaller increases in channel capacity.

As already mentioned, the channel capacity of a communication channel is, by definition, the maximum rate at which information can be transmitted with an arbitrarily low probability of error. Indeed, the recent vociferous scientific debates about the channel capacity of different channels have led to the realisation that all the optical channels discussed are indeed different: they depend on fibre and amplifier type, span length and the in-span compensation scheme – such as optical phase conjugation or optical regeneration. In the case of the NL channel, the channel properties become power dependent either at the onset of significant nonlinearity (*i.e.,* around the power threshold P_{opt}) or in the high power regime. Each of these different channels will have its own channel capacity and comparisons are not obvious, and must be made with

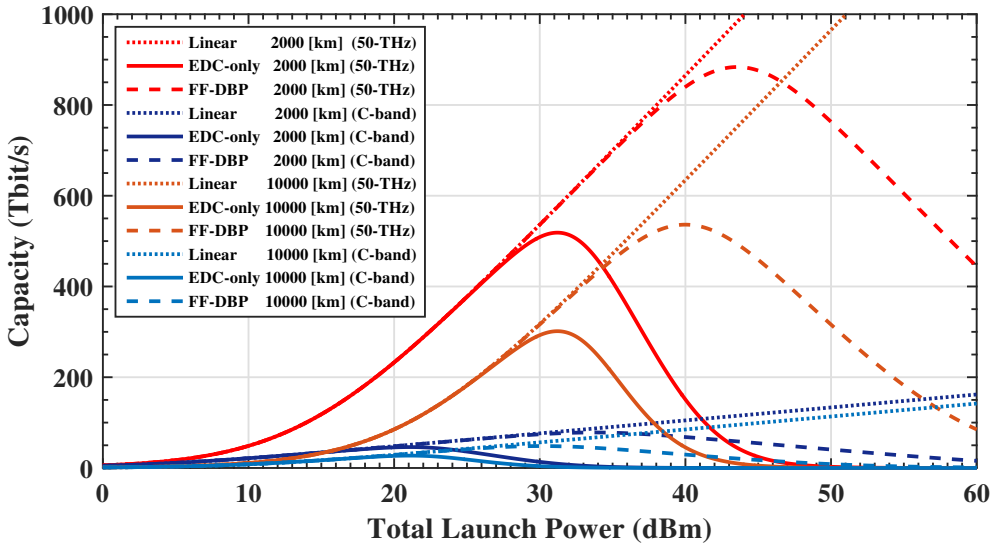

Figure 4.7: Capacity estimate versus total launched power for C-band and 50 THz at 2000 km (25 spans of 80 km) and 10000 km (125 spans of 80 km).

utmost care.

Assuming an AWGN channel, one could use the GN-model to estimate the required optical power to ensure approximately a factor of 10 increase in capacity compared to the current record. It is possible to explore the limits of what is possible through nonlinearity compensation. Using somewhat unrealistic, simplified, assumptions (*e.g.,* wavelength-independent dispersion and loss coefficients), it is possible to explore what capacity gains might be achievable through (i) complete nonlinearity compensation and (ii) an increase in usable fibre bandwidth beyond the current EDFA bandwidth limit of approximately 35 nm (4.3 THz) to 50 THz (*i.e.,* the full standard single-mode fibre bandwidth), approximately an order of magnitude increase. The parameters which have been assumed for the calculations include: 2 polarisations, 50 THz bandwidth (50 Gbaud × 1000 channels), C-band (4.3 THz, 50 GBaud×86 channels), lumped amplification with NF = 3 dB, group velocity dispersion $\beta_2 = -21.7\,\mathrm{ps}^2\,\mathrm{km}^{-1}$, zero PMD, NL coefficient $\gamma = 1.2\,\mathrm{W}^{-1}\,\mathrm{km}^{-1}$, attenuation $\alpha = 0.2\,\mathrm{dB}\,\mathrm{km}^{-1}$ and 80 km per span.

The obtained results are illustrated in Fig. 4.7, where capacity is plotted against power for two different bandwidths and distances (2000 km – a long-haul terrestrial link; and 10000 km – equivalent to a transoceanic system). It can be seen from Fig. 4.7, for a spectral efficiency of approximately 10 bit/s/Hz per polarisation, it would be possible to transmit 1 Pbit/s in the linear regime with quantum noise-limited amplifiers at the launched power of

45 dBm (approximately 30 W total launched power). This would exceed the current record for maximum system capacity of 100 Tbit/s by approximately a factor of 10. Taking nonlinearity of the channel into account leads to a reduction in capacity of approximately 500 Tbit/s. Much of this capacity can be recovered by FF DBP, which increases capacity to 850 Tbit/s. Without relying on the use of spatial modes, a factor of greater than 10 is difficult to envisage at present because of the difficulty of increasing transmitted capacity by purely increasing the cardinality of modulation format. For instance, DP 64-QAM requires 26 distinct amplitude-phase levels per polarisation, yielding 12 bit/s/Hz maximum spectral efficiency (or 6 bit/s/Hz per polarisation). These simple calculations allow us to explore the price one would have to pay in terms of SNR required to achieve these gains, and the associated optical power. Note, again, that there is an assumption of the fundamental limit in amplifier noise figure of 3 dB.

Utilising the above-mentioned GN-model, a three-dimensional sweep of signal launched power, capacity and transmission distance is plotted in Fig. 4.8, showing the relationships among these parameters. Fig. 4.9 shows the capacity obtained when the launched power is optimised for each distance and both the EDC only and FF DBP. For the EDC only case, the optimum launched power is independent of transmission distance and is given by

$$P_{\mathrm{opt}} = \sqrt[3]{\frac{\sigma^2_{\mathrm{ASE}}}{2\eta}}\,. \tag{4.13}$$

For the FF DBP case, the optimum launched power depends on transmission distance (number of spans), and can be approximated as[3]

$$P_{\mathrm{opt}} = \sqrt{\frac{2}{3\,(N_{\mathrm{s}}-1)\,\eta}}\,, \quad N_{\mathrm{s}} \geq 2\,. \tag{4.14}$$

As we can see, in the case of FF DBP, the optimum launched power is distance dependent irrespective of the ASE noise power.

Considering the results in Fig. 4.9, it is possible to use empirical data to estimate the feasibility of achieving, or indeed of going beyond, the above scenario. To encode 850 Tbit/s in 50 THz optical bandwidth requires 8.5 bit/symbol per polarisation at the Nyquist limit. To date, the record achieved spectral efficiency for gigabit-class fibre-optic communications is 7.65

[3]Here we neglect the second-order SNI (see Sec. 3.4), and the coherence factor ϵ is set to zero.

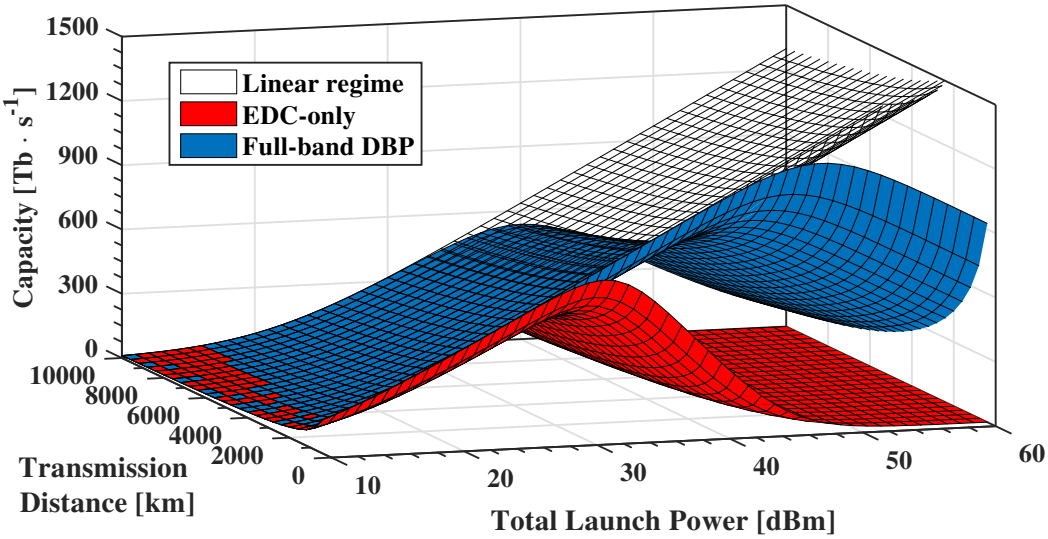

Figure 4.8: Estimated capacity versus transmission distance and total launched power at 50 THz bandwidth

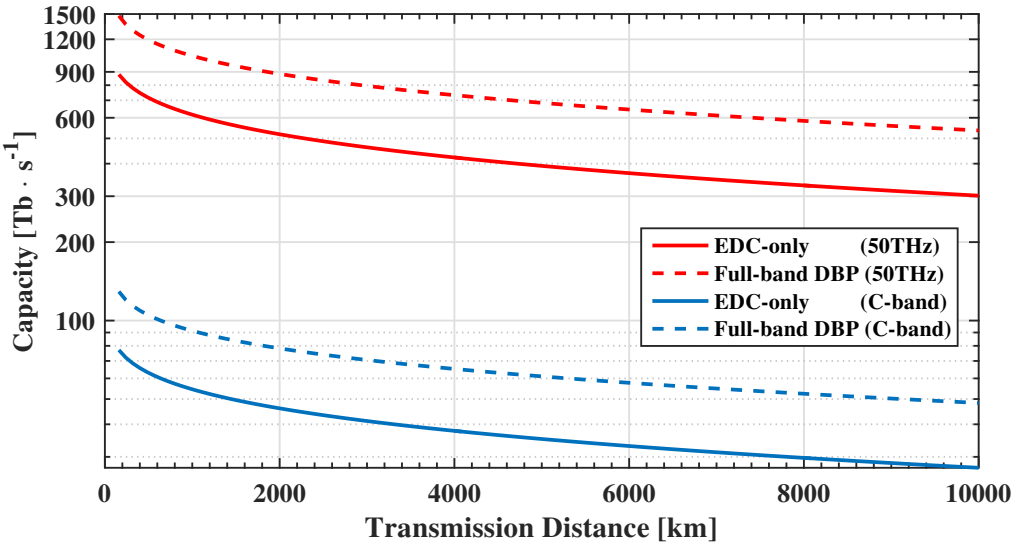

Figure 4.9: Estimated capacity) capacity versus distance for optimum launch power for both 50 THz bandwidth and *C*-band.

bit/symbol per polarisation (considering FEC overhead, which can be as high as 100%, depending on the combination of nonlinearity, transmitted format and distance) [150], encoded using DP 2048-QAM. Note that, although this does not achieve the required information per symbol, higher order constellations could bridge the gap in this case. The limiting factors in increasing the order of QAM beyond this level include: the availability of high-speed, high-resolution DACs and ADCs, laser line-width and the accuracy of the DSP

implementation. All of these factors will in some way limit the achievable SNR; however, current hardware limitations are unlikely to persist, meaning that the above scenario may, at some point, be experimentally demonstrated.

In order to explore the possibilities beyond the 850 Tbit/s transmission considered above, it is instructive to look at (traditional) coaxial communications, where high-order QAM signal generation and detection is more advanced. The current Data Over Cable Service Interface Specification (DOCSIS) 3.1 standard documents the required capability of transmitting 4096-QAM, and the possibility of transmitting 16384-QAM [151], which encodes 14 bit/symbol. On the horizon, there is the possibility of transmitting 65536-QAM within DOCSIS; however the SNR requirements for this format are in excess of 60 dB. Considering, again, the 50 THz bandwidth used in the above transmission scenario, but assuming that the SNR is, now, sufficient to transmit DP 65536-QAM, the maximum capacity would be limited to 1.6 Pbit/s, which, in spite of the staggering SNR requirement, is less than a factor of two beyond what could be achieved using the already-demonstrated DP 2048-QAM. Thus, for future ultra-high-capacity systems, there is a need to unlock ever wider optical fibre bandwidth for transmission. This would include, for example, development of new ultra-wideband amplifiers, sources, detectors and fibre designs combined with space-division multiplexing. The latter may include multiple fibres, multiple cores within a single fibre or multiple modes within a single fibre. Metamaterials could offer all-optical nonlinearity compensation through the use of negative n_2 materials [152]. There remain many challenges to overcome the NL limits of optical fibre communications, to show that already realisable capacities can be achieved over ever-greater distances. Finally and optimistically, even relatively small increases in the achievable transmission rates may be converted into large gains in overall network capacity for networks operated in the NL regime. This is a relatively unexplored area and requires the evaluation of overall network throughput, taking channel properties into account. Given that the definition of network throughput and/or capacity is itself under debate – how much more so techniques for quantifying it.

4.4 Summary

The AIRs of Nyquist-spaced WDM optical communication systems were analytically estimated using ultra-wide bandwidth first-order perturbation analysis for ∼100 nm (∼12.5 THz) backward-pumped Raman amplification and ∼40 nm (∼5 THz) of EDFA amplification over a SSMF, with the ap-

plications of EDC and FF DBP. In order to to demonstrate the possibility of approaching the fundamental limits of the two most widely considered amplification schemes, higher-order modulation formats, and probabilistically-shaped signal constellations have been considered. The nonlinearities within each channel are individually evaluated, and the AIRs were investigated with the applications of EDC and FF DBP. In addition, it is shown that the use of PS, together with FF DBP, can further improve the AIRs and approach the confines imposed by the interactions between the transmitted signal and ASE noise.

Additionally, it is shown that using FF DBP and PS, a DP 256-QAM system can achieve the same AIRs as the DP 1024-QAM system, when the transmission distance exceeds 3200 km in the EDFA scheme and 6000 km in Raman amplification scheme.

Chapter 5

Capacity Bounds of Fibre Channel

In previous chapters, we considered the performance (bounds on the AIR) of ideal 2000 km and 10000 km fibre-optic communication systems, which are referred to as dispersion-unmanaged communication systems, in which the chromatic dispersion is assumed to be compensated in the electrical domain exploiting well-developed EDC technique. In those systems, due to the influence of the accumulated chromatic dispersion throughout the propagation process, the distribution of the propagated optical signal can be quite accurately approximated by the Gaussian distribution. This chapter is dedicated to a new fibre-optic communication scheme in which information is assumed to be embedded into soliton amplitudes (a train of solitons in time domain), and then detected via the so-called forward NFT. These proposed information-theoretic discrete-time channel model takes into account both chromatic dispersion and Kerr nonlinearity. Indeed, the proposed fibre-optical communication scheme requires neither chromatic dispersion nor nonlinearity compensation techniques. It is also worth noting that the channel law of the proposed theoretical channel model for the proposed communication scheme is no longer a conditional Gaussian distribution function.

5.1 Literature review

The fibre channel model in optical communication that includes the effects of attenuation, chromatic dispersion, and nonlinearity for one state of polarisation is the scalar NLSE [135, 153, 83, 154], whereas for two states of polarisation, it is the vectorial ME [155, Eq. (1.26)], [153, Sec. 10.3.1]. The ME describes the propagation of the optical electric field employing the polarisation division multiplexing technique. Therefore, the ME is the generalisation of

the popular scalar NLSE. As a matter of fact, in both models, the evolution of the optical field along the fibre is represented by a *generalised* inhomogeneous partial differential equation including the stochastic term on the right-hand side, which is usually modelled as a complex additive Gaussian noise. Note that, the accumulated NL interaction between the signal and the noise makes the analysis of the resulting channel model a very difficult problem. As recently discussed in, *e.g.,* [23, Sec. 1], [24, 25], exact channel capacity results for fibre optical systems are scare, and many aspects related to this problem still remain open.

Until recently, the common belief among some researchers in the field of optical communication was that optical nonlinearity is always a nuisance that necessarily degrades the fibre-optic communication system performance. That, in turn, led to the assumption that the throughput of an optical channel had a peaky behaviour when plotted as a function of the launched power, *i.e.,* the maximum is reached at a finite threshold power[1]. Partially motivated by the idea of improving the data rates in optical fibre links, a multitude of nonlinearity mitigation methods have been proposed over recent years to mitigate nonlinearity-induced distortions, each resulting in different discrete-time channel models. This includes receiver-based digital signal processing [157], digital backpropagation [66], optical phase conjugation [158], twin-waves phase conjugation [159], *etc.* Recently, a paradigm-shifting approach for overcoming the detrimental impact of nonlinearity has been receiving increased attention. This approach relies on the fact that, in the absence of both losses and noise the partial differential equations describing the propagation of optical waveforms through optical fibre belong to the class of *exactly integrable* (*i.e.,* completely solvable) evolutionary equations [44, 160]. The loss in such optical links is usually assumed to be compensated by using either lumped or distributed amplification, particularly, the IDRA scheme [161, 162] resulting in the lossless NLSE [135]. The optical signal will be nonetheless perturbed by ASE noise as well as the noise due to a signal-noise beating.

One of the main consequence of the property of integrability of noiseless and lossless propagation equations is that the optical signal evolution can be essentially represented using NL normal modes. Whilst the pulse propagation in the ME and NLSE is, in fact, strictly NL, the evolution of these NL modes in the so-called NL spectral domain is essentially linear [163, 164]. The decomposition of the waveforms into the NL modes (and the reciprocal operation)

[1]However, non-decaying bounds can be found in the literature, *e.g.,* in [23] and [156] (lower bounds) and [148] and [149] (upper bounds).

is often referred to as NFT, because of its similarity with the application of the conventional Fourier decomposition in linear systems [165][2]. The linear propagation of the NL modes implies that NL cross-talk in the NFT domain is theoretically absent, an idea exploited in the so-called *nonlinear frequency division multiplexing* [163, 166]. Using this approach, the NL interference can be greatly suppressed by assigning users different ranges in the NL spectrum, instead of multiplexing them via the conventional Fourier domain.

Integrability (and the general ideas based around NFT) has also given rise to several nonlinearity compensation, transmission, and coding schemes. These can be seen as a generalisation of soliton-based communications, which follow the pioneering work by Hasegawa and Nyu, in which the concept of *eigenvalue communications* was introduced [167]. However, only the discrete eigenvalues associated with the solitonic degrees of freedom emerging from the NFT signal decomposition were used for communication purpose. The development of efficient and numerically stable algorithms has also attracted a lot of attention [168]. Moreover, there have been a number of experimental demonstrations and assessments for different NFT-based systems.

Two NL spectra (types of NL modes) exist in both the NLSE and the ME. The first one is the so-called *continuous spectrum*, which is the exact NL analogue of the familiar linear Fourier transform (FT), inasmuch as its evolution in an optical fibre is precisely equivalent to that of the linear spectrum under the action of the chromatic dispersion and the energy contained in the continuous spectrum is related to that in the time domain by a modified Parseval equality [165, 169]. The unique feature of the NFT is, however, that apart from the continuous spectrum, it can support a set of discrete eigenvalues (the non-dispersive part of the solution), *i.e.,* the *discrete spectrum*, having no analog in linear optics. In the time domain, these eigenvalues correspond to stable localised multi-soliton waveforms immune entirely to perturbations [153]. The spectral efficiency of the multiple-eigenvalue encoding schemes is an area actively explored at the moment [170, 171, 172]. Multi-soliton transmission has also received increased attention in recent years, see, *e.g.,* [173] and [174] and corresponding references therein. Finding the capacity of the multi-eigenvalue-based systems in the presence of in-line noise that breaks down integrability still remains an open research problem. If only a single eigenvalue per time slot is used, the problem is equivalent to a well-known

[2]In mathematics and physics literature, the name *inverse scattering transform* method for the NFT is more commonly used.

time-domain amplitude-modulated soliton transmission system[3]. In this chapter, we consider this simple set-up, where a single eigenvalue is transmitted in every time slot. The obtained results are straightforwardly applicable not only to classical soliton communication systems, but also to the novel area of the eigenvalue communications.

Although the set-up we consider in this chapter is one of the simplest ones, its channel capacity is still unknown. In particular, previous results include those by Meron *et al.* [175], who recognised that MI in a NL integrable channel can (and should) be evaluated through the statistics of the nonlinear spectrum, *i.e.,* via the channel defined in the NFT domain. Using a Gaussian scalar model for the amplitude evolution with in-line noise, a lower bound on the MI and capacity of a single-soliton transmission system was presented. The case of two and more solitons per one time slot was also analysed, where data rate gains of the continuous soliton modulation versus an on-off keying (OOK) system were also shown. The first attempt to analyse a bit error ratio (BER) for the case of two interacting solitons has been made in [176]. The derivations presented there, however, cannot be implemented straightforwardly for information theoretic analysis. Yousefi and Kschischang [170] addressed the question of achievable spectral efficiency for single- and multi-eigenvalue transmission systems using a Gaussian model for the nonlinear spectrum evolution. Some results on the continuous spectrum modulation were also presented. Later, in [171], the spectral efficiency of a multi-eigenvalue transmission system was studied in more detail. In [172], the same problem was studied by considering the correlation functions of the spectral data obtained in the quasi-classical limit of large number of eigenvalues. Achievable information rates for multi-eigenvalue transmission systems utilising all four degrees of freedom of each scalar soliton in NSE were analytically obtained in [177]. These results were obtained within the framework of a Gaussian noise model provided in [170] and [178] (non-Gaussian models have been presented in [179] and [180]) and assuming a continuous uniform input distribution subject to peak power constraints. The spectral efficiency for the NFT continuous spectrum modulation was considered in [181, 182, 183]. Periodic NFT methods have recently been investigated in [184].

[3]Since the imaginary part of a single discrete eigenvalue is proportional to the soliton amplitude.

5.2 Normalisation and solitonic units

The propagation of light in optical fibres in the presence of ASE noise can be described by a stochastic partial differential equation, which captures the effects of chromatic dispersion, polarisation mode dispersion, the optical Kerr effect, and the generation of ASE noise from the optical amplification process. Throughout this chapter we assume that the fibre loss is continuously compensated along the fibre by means of IDRA [161, 162]. We consider the propagation of a slowly varying 2-component envelope $\mathbf{Q}(\ell,\tau)=[Q_1(\ell,\tau),Q_2(\ell,\tau)]\in\mathbb{C}^2$ over a NL birefringent optical fibre, where ℓ and τ represent propagation distance and time, respectively. Our model also includes the 2-component ASE noise $\mathbf{N}(\ell,\tau)=[N_1(\ell,\tau),N_2(\ell,\tau)]$ due to the DRA. Here we also assume a uniform change of polarised state of the Poincaré sphere [52].

The resulting lossless generalised *vector* ME is then given by [155, Eq. (1.26)], [153, Sec. 10.3.1], [185], [186] (see also Eq. (2.52))[4]

$$\imath\frac{\partial\mathbf{Q}(\ell,\tau)}{\partial\ell}-\frac{\beta_2}{2}\frac{\partial^2\mathbf{Q}(\ell,\tau)}{\partial\tau^2}+\frac{8\gamma}{9}\left\langle\mathbf{Q}(\ell,\tau),\mathbf{Q}^\dagger(\ell,\tau)\right\rangle\mathbf{Q}(\ell,\tau)=\mathbf{N}(\ell,\tau)\,,\qquad(5.1)$$

where the retarded time τ is measured in the reference frame moving with the optical pulse average group velocity, $\mathbf{Q}\equiv\mathbf{Q}(\ell,\tau)$ represents the slowly varying 2-component envelope of electric field, β_2 is the group velocity dispersion coefficient characterising the chromatic dispersion, and γ is the fibre nonlinearity coefficient. The pre-factor 8/9 in Eq. (5.1) comes from the averaging of the fast polarisation rotation, detailed explanation and derivation are provided in [153, Sec. 10.3.1], [52]. For simplicity we shall further work with the *effective* averaged NL coefficient $\gamma^*\triangleq 8\gamma/9$ when addressing the ME. In the case of a single polarisation state, the propagation equation above reduces to the lossless generalised *scalar* NLSE (see Eq. (2.46))

$$\imath\frac{\partial Q(\ell,\tau)}{\partial\ell}-\frac{\beta_2}{2}\frac{\partial^2Q(\ell,\tau)}{\partial\tau^2}+\gamma\left|Q(\ell,\tau)\right|^2Q(\ell,\tau)=N(\ell,\tau)\,.\qquad(5.2)$$

In this chapter, we consider the case of anomalous dispersion ($\beta_2<0$), *i.e.,* the *focusing* case, which is associated with "anomalous dispersion" in optical fibre In this case, both the ME in Eq. (5.1) and the NLSE in Eq. (5.2) permit bright soliton solutions ("particle-like waves"), which will be discussed in more detail in Sec. 5.4.

It is customary to re-scale Eq. (5.1) to dimensionless units. We shall use

[4]Throughout this chapter ℓ and τ denote the real-world units, while z and t are the normalised dimensionless distance and time, respectively.

the following normalisation: The power will be measured in units of $P_0 = 1\,\mathrm{mW}$ since it is a typical power level used in optical communications. The normalised (dimensionless) field then becomes $\mathbf{q} = \mathbf{Q}/\sqrt{P_0}$. For the distance and time, we define the dimensionless variables $z = \ell/\ell_0$ and $t = \tau/\tau_0$, where

$$\begin{aligned} \ell_0 &= (\gamma^* P_0)^{-1}, \\ \tau_0 &= \sqrt{\ell_0 \, |\beta_2|} = \sqrt{\frac{|\beta_2|}{\gamma^* P_0}}. \end{aligned} \tag{5.3}$$

For the scalar case (5.2), we use the same normalisation but we replace γ^* by γ. Then, the resulting ME finally reads

$$\imath \frac{\partial \mathbf{q}(z,t)}{\partial z} + \frac{1}{2} \frac{\partial^2 \mathbf{q}(z,t)}{\partial t} + \Big\langle \mathbf{q}(z,t), \mathbf{q}^\dagger(z,t) \Big\rangle \, \mathbf{q}(z,t) = \mathbf{n}(z,t), \tag{5.4}$$

whilst the NLSE becomes

$$\imath \frac{\partial q(z,t)}{\partial z} + \frac{1}{2} \frac{\partial^2 q(z,t)}{\partial t^2} + |q(z,t)|^2 \, q(z,t) = n(z,t). \tag{5.5}$$

The ASE noise $\mathbf{n}(z,t) = [n_1(z,t), n_2(z,t)]$ in (5.4) is a normalised version of $N(\ell,\tau)$, and is assumed to have the following correlation properties

$$\begin{aligned} &\mathbb{E}\,[n_i(z,t)] = \mathbb{E}\left[n_i(z,t)\, n_j(z',t')\right] = 0, \\ &\mathbb{E}\left[n_i(z,t)\, \bar{n}_j(z',t')\right] = D\, \delta_{ij}\, \delta\left(z - z'\right) \delta\left(t - t'\right), \end{aligned} \tag{5.6}$$

with $i,j \in \{1,2\}$, with δ_{ij} being a Kronecker symbol, $\mathbb{E}\,[\cdot]$ is the mathematical expectation operator, and $\delta\,(\cdot)$ is the Dirac delta function. The correlation properties (5.6) mean that each noise component $n_i(z,t)$ is assumed to be a zero-mean, independent, white circular Gaussian noise. The scalar case follows by considering a single noise component only. The noise intensity D in (5.6) is (in dimensionless units)

$$D = \sigma_0^2 \frac{\ell_0}{P_0 \tau_0} = \frac{\sigma_0^2}{\sqrt{\gamma^* \, |\beta_2| \, P_0^3}}, \tag{5.7}$$

where σ_0^2 is the spectral density of the noise, with real world units $[\mathrm{W}/(\mathrm{km} \cdot \mathrm{Hz})]$. For IDRA, this σ_0^2 can be expressed through the optical fiber and transmission system parameters as follows: $\sigma_0^2 = \alpha_{\mathrm{fiber}} K_T \cdot h\nu_0$, where α_{fiber} is the fibre attenuation coefficient, $h\nu_0$ is the average photon energy, K_T is a temperature-dependent phonon occupancy factor Eq. (3.20).

Note that, from now on, all the quantities in this chapter are given in normalised units unless specified otherwise. Furthermore, we define the continuous time channel as the one defined by the normalised ME and the NLSE. This is shown schematically in the inner part of Fig. 5.1, where the transmitted and received waveforms are $\mathbf{x}(t) \equiv \mathbf{q}(0,t)$ and $\mathbf{y}(t) \equiv \mathbf{q}(Z,t)$, respectively, where Z is the propagation distance.

5.3 Forward nonlinear Fourier transform

The waveform continuous-time channel defined by the scalar NLSE Eq. (5.5) is fundamentally NL, and thus, the conventional Fourier transform analysis is no longer applicable. Additionally, the connection between the continuous-time physical channel model and the discrete-time information-theoretic channel model cannot be realised via the Nyquist sampling, since the bandwidth of input signal waveform evolves to infinity due to the impact of the NL interference process. This fosters research to seek principally new approaches, which are potentially more natural and suitable for inherently NL optical communication systems.

In the late 1960's, Gardner *et al.* [187] first applied a mathematical method, which is nowadays widely recognised in physical and mathematical communities as the *inverse scattering transform*. This approach could be deemed a generalisation of the conventional Fourier transform method that allowed them to solve, in the closed form, the initial-value problem for the NL partial differential equation known as the Korteweg-de Vries equation. The solution in the closed-form was obtained similarly to the application of standard Fourier decomposition method for linear partial differential equations (*e.g.,* the heat equation)[5]. After that discovery, other NL generalisations of the Fourier transform were found to solve other important NL systems, in particular – the NLSE, where the explicit form of the respective operations were first presented in the seminal paper by Zakharov and Shabat in 1972 [44]. These generalisations of the Fourier transform to the case of NL partial differential equations are now often referred to as the NFT method in the signal-processing literature. The equations and associated initial-value problems, which can be solved by means of the NFT are called *integrable*, such that the integrability and the NFT solution are often mentioned in the same context. Thus, the

[5]In 1968, Peter Lax pointed out that any NL wave equation can be integrable via the inverse scattering transform method if and only if there exist the so-called *compatibility* condition between two linear operators, which are now commonly referred to as the *Lax pair* [188].

applicability of the NFT method to the lossless and noiseless NLSE implies that the NLSE belongs to the class of completely integrable partial differential equations[6]. The NLSE for the anomalous dispersion ($\beta_2 < 0$) case (*i.e.,* the focusing case) possesses both localised particle-like solitary waves (optical solitons) and (quasi)-linear radiation waves, having no analogies in linear systems. The notable property of solitons is their aptitude to pass through each other without any changing of shape and speed, acquiring merely a phase change. It is precisely this property that caused the first surge of attention to them in the optical communications community as they were thought to be good candidates for robustly carrying the encoded information over a NL fibre.

Note that, in this section, we provide a brief description of the forward NFT only, since the forward NFT block is employed as the receiver in our proposed fibre-optic communication scheme (see Fig. 5.1). The NFT decomposition of a signal $q(z,t)$ associated with the scalar NLSE Eq. (5.5) is determined via the spectral analysis of the linear Lax operator $\hat{\mathcal{L}}$, *i.e.,*

$$\hat{\mathcal{L}} = \begin{pmatrix} \imath \frac{\partial}{\partial t} & q(z,t) \\ -q^*(z,t) & -\imath \frac{\partial}{\partial t} \end{pmatrix}, \tag{5.8}$$

where the input signal waveform $q(z,t)$ enters as an effective potential. The spectrum of $\hat{\mathcal{L}}$ can be found as a solution of the following eigenvalue problem, which is written down as an auxiliary two-dimensional function $\boldsymbol{\psi} \equiv \boldsymbol{\psi}(z,t;\lambda) = [\psi_1(z,t;\lambda)\,, \psi_2(z,t;\lambda)]^{\mathrm{T}}$, *i.e.,*

$$\hat{\mathcal{L}}\boldsymbol{\psi} = \lambda\boldsymbol{\psi}\,, \tag{5.9}$$

where the (generally complex-valued) scattering parameter $\lambda \in \mathbb{C}$ can be understood as the NL analogue of the conventional frequency.[7] As λ takes complex values for solitons, we can write it as,

$$\lambda = \zeta + \imath\xi\,, \tag{5.10}$$

where $\zeta = \mathrm{Re}[\lambda]$ and $\xi = \mathrm{Im}[\lambda]$ are the real and imaginary NL frequencies, respectively. The spectrum of the Zakharov-Shabat operator $\hat{\mathcal{L}}$ is found to be *invariant* with z-propagation, as long as $q(z,t)$ evolves according to the noiseless and lossless NLSE Eq. (5.5); in other words, the complex eigenvalues

[6]The general problem that allows derivation of a class of integrable equations known as the Ablowitz-Kaup-Newell-Segur auxiliary problem [189].

[7]Here the term "frequency" was fairly justified since it was shown that the NFT continuous spectrum approaches the conventional linear Fourier spectrum with asymptotically decreasing the signal power. Indeed, the eigenvalues λ are connected to the Fourier frequencies ν as follows $\lambda = -\pi\nu$ [190].

of this spectral problem are the integrals of motion, *i.e.,* they remain the same throughout the propagation. This property constitutes the main idea of NFT-based communication, that is, to employ these invariant eigenvalues to encode and then carry information with no distortions from the source to the destination over a NL fibre. It can be shown that the NFT spectrum, *i.e.,* the set of data emerging from the NFT decomposition of the signal waveform evolves in a purely linear decoupled manner, whilst the waveform signal itself experiences a highly complicated NL evolution in the *true* space-time domain.

Eq. (5.9) is equivalent to the following system of linear temporal and spacial differential equations

$$\frac{\partial \boldsymbol{\psi}}{\partial t} = \hat{\mathcal{U}}\,\boldsymbol{\psi}\,, \tag{5.11}$$

$$\frac{\partial \boldsymbol{\psi}}{\partial z} = \hat{\mathcal{V}}\,\boldsymbol{\psi}\,, \tag{5.12}$$

where operators $\hat{\mathcal{U}}$ and $\hat{\mathcal{V}}$ are known as the Lax pair. z-invariance of the eigenvalue, *i.e.,* $\frac{\partial \psi}{\partial z} = 0$ gives rise to the following compatibility condition, also known as the *zero-curvature* equation,

$$\frac{\partial \hat{\mathcal{U}}}{\partial z} - \frac{\partial \hat{\mathcal{V}}}{\partial t} + \left[\hat{\mathcal{U}}, \hat{\mathcal{V}}\right] = 0\,, \tag{5.13}$$

with $\left[\hat{\mathcal{U}}, \hat{\mathcal{V}}\right] \triangleq \hat{\mathcal{U}}\hat{\mathcal{V}} - \hat{\mathcal{V}}\hat{\mathcal{U}}$ being the commutator. The Lax pair for the lossless and noiseless NLSE Eq. (5.5) was discovered by Zakharov and Shabat

$$\hat{\mathcal{U}} = \begin{pmatrix} -\imath\lambda & \imath q(z,t) \\ \imath q^*(z,t) & \imath\lambda \end{pmatrix}, \tag{5.14}$$

$$\hat{\mathcal{V}} = \begin{pmatrix} \imath\lambda^2 + \frac{\imath}{2}\left|q(z,t)\right|^2 & \imath\lambda q(z,t) - \frac{1}{2}\frac{\partial q(z,t)}{\partial t} \\ \imath\lambda q^*(z,t) - \frac{1}{2}\frac{\partial q^*(z,t)}{\partial t} & \imath\lambda^2 - \frac{\imath}{2}\left|q(z,t)\right|^2 \end{pmatrix}. \tag{5.15}$$

Eqs (5.11), (5.12) along with operators Eqs (5.14), (5.15) are the so-called Zakharov-Shabat equations, which are the coupled linear ordinary differential equations. Thus, this procedure reduces the integration of NLSE to the integration of the system of linear equations.

Suppose that the optical signal $q(t)$ is absolutely integrable function [191, 192], *i.e.,*

$$\int_{-\infty}^{\infty} dt\,\left|q(z,t)\right| < \infty\,. \tag{5.16}$$

In order to find the NFT spectrum associated with the profile $q(z)$, we fix the

"initial" condition for the incident wave scattered by the potential $q(z,t)$, at the trailing of our localised signal, *i.e.,*

$$\lim_{t\to-\infty} \mathbf{\Psi}(z,t;\lambda) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \exp(-\imath\lambda t). \tag{5.17}$$

where $\mathbf{\Psi}(z,t;\lambda) = [\,\Psi_1(z,t;\lambda)\,,\,\Psi_2(z,t;\lambda)\,]^{\mathrm{T}}$ is called the *Jost function*. Utilising Eq. (5.17), we can now determine the following *Jost scattering coefficients*

$$a(\lambda) = \lim_{t\to\infty} \Psi_1(z,t;\lambda)\exp(\imath\lambda t)\,, \tag{5.18}$$

$$b(\lambda) = \lim_{t\to\infty} \Psi_2(z,t;\lambda)\exp(-\imath\lambda t)\,. \tag{5.19}$$

The NFT acts on the waveform $q(z,t)$ function, and the result consists of *continuous* $\mathscr{Q}_{\mathrm{c}}(\zeta)$ and *discrete* $\mathscr{Q}_{\mathrm{d}}(\lambda_j)$ spectra. The continuous part represents the *right* reflection coefficient [193] associated with Eq. (5.11), which is given by the following spectral function *i.e.,*

$$\mathscr{Q}_{\mathrm{c}}(\zeta) \triangleq \frac{b(\zeta)}{a(\zeta)} = \lim_{t\to\infty} \frac{\Psi_2(z,t;\zeta)}{\Psi_1(z,t;\zeta)} \exp(-2\imath\zeta t)\,, \quad \zeta\in\mathbb{R}, \tag{5.20}$$

where $\zeta\in\mathbb{R}$ is the real NL frequency. The discrete part represents the so-called norming constants defined as

$$\mathscr{Q}_{\mathrm{d}}(\lambda_j) \triangleq \frac{b(\lambda_j)}{\left.\frac{da(\lambda_j)}{d\lambda}\right|_{\lambda=\lambda_j}}, \quad \lambda_j\in\mathbb{C}^+, \quad j=1,\ldots,N, \tag{5.21}$$

with N being the number of eigenvalues. The set $\{\lambda_j\}_{j=1}^{N}$ denotes the zeros of $a(\lambda)$.

To summarise, the forward NFT operation corresponds mapping of the input optical waveform $q(t)\equiv q(0,t)$ onto the set of scattering data (NFT-spectrum) Σ, *i.e.,*

$$\begin{aligned} &q(t) \xrightarrow{\text{forward NFT}} \Sigma, \\ &\Sigma \triangleq \left\{ \mathscr{Q}_{\mathrm{c}}(\lambda)\,,\,\zeta\in\mathbb{R};\, \left\{ \{\lambda_j\}_{j=1}^{N}\,,\, \mathscr{Q}_{\mathrm{d}}(\lambda_j)\,,\, \lambda_j\in\mathbb{C}^+ \right\} \right\}. \end{aligned} \tag{5.22}$$

The reciprocal operation corresponds to the inverse NFT, which is, however, not needed for our proposed communication system. It is important that the discrete NFT spectrum $\mathscr{Q}_{\mathrm{d}}(\lambda_j)$ as well as the isolated points $\{\lambda_j\}_{j=1}^{N}$ are associated with the solitonic (non-dispersive) degrees of freedom, whereas the

continuous $\mathscr{Q}_{\mathrm{c}}(\zeta)$ spectrum is related to the radiative (dispersive) waves. In particular, in the case of the scalar NLSE, the optical soliton has four degrees of freedom (details will be discussed in the next Sec. 5.4): the soliton frequency (sometimes called the soliton velocity) Ω and the soliton amplitude A in Eq. (5.24) are connected to the real and imaginary parts of eigenvalues λ in Eq. (5.10) as follows

$$\zeta = \frac{1}{2}\Omega\,, \quad \xi = \frac{1}{2}A\,, \tag{5.23}$$

and the discrete spectral amplitudes $\mathscr{Q}_{\mathrm{d}}(\lambda_j)$ determines the phase ϕ and the centre position of the soliton T in Eq. (5.24).

5.4 Bright solitons solution

The optical solitons are localised particle-like waves, which, in essence, always remain unchanged after interactions with other solitons of the same type [194]. Thus, they can often be employed to encode and transmit information in optical communication systems. It has been established that the unperturbed noiseless scalar NLSE possesses a special class of solutions, the so-called fundamental *bright solitons*. It may be worth noting that fundamental solitons are "bright" only for the focusing case we consider in this chapter, *i.e.,* for the anomalous dispersion case. The general form of the 1-soliton solution of (5.5) with $n(z,t) = 0$ is given by the following closed-form expression

$$q(z,t) = A(z)\,\mathrm{sech}\Big\{A(z)\,[t - T(z)]\Big\}\exp\Big[-\imath\,\Omega(z)\,t + \imath\,\phi(z)\Big]\,, \tag{5.24}$$

where $\mathrm{sech}(\cdot)$ is the hyperbolic sine, $A(z)$, $\Omega(z)$, $\phi(z)$, $T(z)$ are the soliton degrees of freedom: soliton amplitude, soliton velocity (frequency), soliton phase, and centre-of-mass position of the soliton, respectively. The dependence of these four parameters in Eq. (5.24) on propagation distance is given by

$$A(z) = A_0\,, \tag{5.25}$$

$$\Omega(z) = \Omega_0\,, \tag{5.26}$$

$$\phi(z) = \phi_0 + \frac{A_0^2 - \Omega_0^2}{2}\,z\,, \tag{5.27}$$

$$T(z) = T_0 - \Omega_0\,z\,, \tag{5.28}$$

where A_0, Ω_0, ϕ_0, T_0 are the initial values of scalar soliton parameters at distance $z = 0$. Without loss of generality, from now we assume that the initial soliton frequency, phase and centre-of-mass position are set to zero. Thus, we

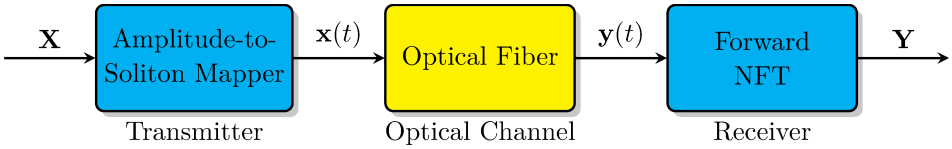

Figure 5.1: System model under consideration. The symbols $\mathbf{X} = [X_1, X_2, X_3, ...]$ are converted to amplitudes, and then mapped to a waveform $\mathbf{x}(t)$. The noisy received waveform $\mathbf{y}(t)$ is obtained by propagating $\mathbf{x}(t)$ in (5.4). The forward NFT processes the waveform $\mathbf{y}(t)$ symbol-by-symbol, and gives a soft estimate of the transmitted symbols $\mathbf{Y} = [Y_1, Y_2, Y_3, ...]$.

have

$$q(0,t) = A_0 \operatorname{sech}(A_0 t), \tag{5.29}$$

and the unperturbed soliton at the distance $z = L$ is

$$q(Z,t) = A_0 \operatorname{sech}(A_0 t) \exp\left(\frac{\imath A_0^2 Z}{2}\right), \tag{5.30}$$

Likewise, in general, the Manakov fundamental soliton is fully characterised by 6 parameters (4 in the NLSE case): frequency (also having the meaning of velocity in some physical applications), phase, phase mismatch, centre-of-mass position, polarisation angle, and amplitude (the latter is inversely proportional to the width of the soliton). We consider amplitude-modulated solitons, and thus, no information is carried by the other 5 parameters. The initial values of these 5 parameters can therefore be set to arbitrary values. We assume that all of them have been set to zero. For the initial frequency, this can be further motivated to avoid deterministic pulse walk-offs. As for the initial phase, phase mismatch, and centre-of-mass position, as we shall see in the next section, their initial values do not affect the marginal amplitude channel law. Under these assumptions, for the noiseless ME the 2-component fundamental bright soliton solution at $z = 0$ reads as (in normalised units) [186, 185]

$$\mathbf{q}(0,t) = A_0 \left[e^{\imath\varphi_0} \cos\psi_0, e^{-\imath\varphi_0} \sin\psi_0\right] \operatorname{sech}(A_0 t), \tag{5.31}$$

where the first term in square brackets corresponds to the first component $q_1(0,t)$, and the second – to $q_2(0,t)$ (and the same will hold further for other 2-component vectors), A_0 denotes the normalised soliton amplitude, $0 < \psi_0 < \pi/2$ is the polarisation angle, and φ_0 is the phase mismatch between two polarisations. The value of ψ_0 can be used to control how the signal power is split across the two polarisation states.

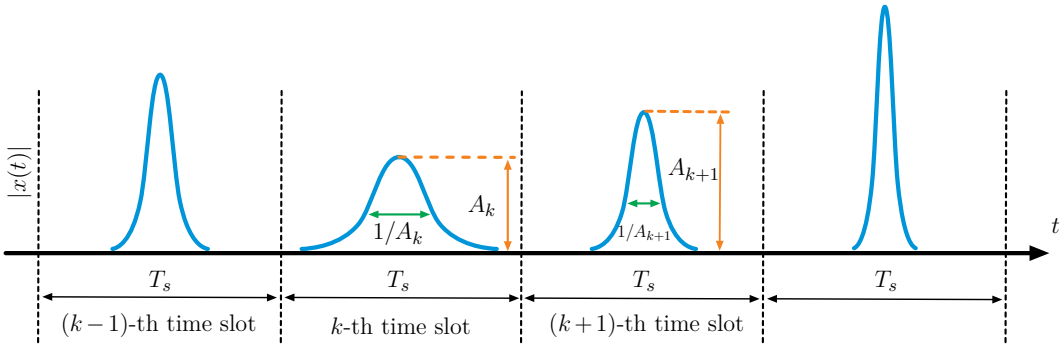

Figure 5.2: Schematic visualisation of the amplitude modulation of soliton sequence (scalar NLSE case).

For any ψ_0, the Manakov soliton solution after propagation over a distance Z with the initial condition given by (5.31), is expressed as

$$\begin{aligned}\mathbf{q}(Z,t) &= A_0 \left[e^{\imath\varphi_0}\cos\psi_0,\, e^{-\imath\varphi_0}\sin\psi_0\right] \operatorname{sech}(A_0 t)\exp\left(\frac{iA_0^2 L}{2}\right)\\ &= \mathbf{q}(0,t)\exp\left(\frac{iA_0^2 Z}{2}\right).\end{aligned} \tag{5.32}$$

Obviously, the soliton solutions Eqs. (5.29) and (5.29) for the NLSE in Eq. (5.5) can be obtained by using $\psi_0 = 0$ and $\varphi_0 = 0$ in Eqs. (5.31)–(5.32)[8].

As shown by Eqs. (5.32) and (5.30), the solitons in Eqs. (5.31) and (5.29) only acquire a phase rotation after propagation. When the noise is no longer assumed to be set to zero, however, these solutions will change. This will be discussed in detail in the next section.

5.5 Discrete-time channel model

We consider a continuous-time input signal $\mathbf{x}(t) = [x_1(t), x_2(t)]$ of the form

$$\mathbf{x}(t) = \sum_{k=1}^{\infty} s_k(t)\,, \tag{5.33}$$

where $s_k(t) = \left[s_{k,1}(t)\,,\, s_{k,2}(t)\right]$ and k is the discrete-time index. Therefore, the pulses $s_k(t)$ are chosen to be

$$s_k(t) = \left[\cos\psi_0\,, \sin\psi_0\right] A_k \operatorname{sech}\left[A_k\,(t - kT_s)\right], \tag{5.34}$$

[8]This corresponds to the case where all the signal power is transmitted in the first polarisation.

where T_s is the symbol period. In principle, it is also possible to encode information by changing the polarisation angle ψ_0 from slot to slot. However, we fix its value to be the same for all the time slots corresponding to a fixed (generally elliptic) degree of polarisation. Thus, the transmitted waveform corresponds to soliton amplitude modulation, which is schematically shown in Fig. 5.2 for the scalar (NLSE) case.

At the transmitter, we assume that symbols X_k are mapped to soliton amplitudes A_k via $A_k = X_k^2$. This normalisation is introduced only to simplify the analytical derivations in this chapter. In order to avoid soliton-to-soliton interactions, we make an assumption that the separation T_s is sufficiently large, *i.e.,* $\exp(-A_k T_s) \ll 1$, $\forall k$. The receiver in Fig. 5.1 is then assumed to process the received waveform during a window of T_s via the forward NFT [160, 195] and returns the amplitude of the received soliton, which is denoted by $R_k = Y_k^2$.

Before proceeding further, it is important to discuss the role of the amplitudes A_k on a potential enhancement of soliton-soliton interactions. The interaction force pre-factor is known to scale as the amplitude cubed [153, Ch. 9.2], [83, Ch. 5.4]. However, the interaction also decays exponentially as follows $\exp(-A_k T_s)$. This exponential decay dominates the interaction, and thus, considering very large amplitudes (or equivalently, very large powers, as we will do later), is, in principle, not an obstacle. At extremely large amplitudes, however, the proposed model used is invalid for different reasons: higher order nonlinearities, which, in fact, are beyond the first-order perturbation analysis, should be taken into account. This includes either stimulated Brilloin scattering (for very large powers) or Raman scattering (for very short pulses, *i.e.,* ultra-wide bandwidths). Studying the impact of these effects is, however, out of the scope of our proposed model.

We would also like to emphasise that for a fixed pulse separation T_s, the channel model we consider is essentially not applicable for low soliton amplitudes. This is mainly due to the following two reasons. The first one is that, for low amplitude solitons, the perturbation theory used to derive the channel law becomes inapplicable as the signal becomes of the same order as noise. Secondly, low amplitude solitons are also very broad, and thus, non-negligible soliton interactions can be potentially generated. These two cases can be overcome if the soliton amplitudes are always forced to be larger than a certain cut-off amplitude $\hat{a}$, which we will now estimate. For the first case (noise-limited), the threshold $\hat{a}_{\text{noise}}$ is proportional to σ_N^2. In the second case (interaction-limited), the threshold is proportional to the symbol rate, *i.e.,* $\hat{a}_{\text{inter}} \propto T_s^{-1}$. This shows that for fixed system parameters, the threshold

$\hat{a} = \max[\hat{a}_{\text{noise}}, \hat{a}_{\text{inter}}]$ is a constant. The implications of this will be discussed at the end of Sec. 5.8.

Another source of corruption for the soliton-based transmission line emanates from the Gordon-Haus (GH) timing jitter [185], defining the variance of the soliton centre position shift as a function of the propagation distance, polarisation angle, and soliton amplitude (here we assume zero phase-mismatch parameter). So, if one neglects the interaction between neighbouring solitons, we then ought to take into account the presence of the GH corruption; the measure of the latter is the standard deviation $\sigma_T^{(\text{GH})}$ (see [185, Eqs. (42), (43)]) associated with the GH effect. For a given propagation distance Z, the inter-soliton separation must fulfil the condition $T_s > \sigma_T^{(\text{GH})}$, and we assume that this condition is always satisfied in our proposed transmission set-up. This condition guarantees that solitons behave as isolated pulses, and thus, there is no inter-symbol interference. A similar discussion for the scalar case can be readily found in [175].

Having defined the transmitter and receiver, we can now define a discrete-time channel model, which encompasses the transmitter, the optical fibre, and the receiver, as shown in Fig. 5.1. Due to the assumption on solitons well-separated in time domain, we do model the channel as memoryless, and thus, from now on we drop the time index k. This memoryless assumption is supported by additional numerical simulations we performed, which are included in Sec. 5.7. Nevertheless, at this point it is hugely important to consider the implications of a potential mismatch between the memoryless assumption of the model and the true channel in the context of channel capacity lower bounds. In particular, if in some regimes (*e.g.,* low power or large transmission distances) the memoryless assumption would not hold, considering a memoryless channel model would result in *approximated* lower bounds on the channel capacity. Provable lower bounds can be obtained by using mismatched decoding theory [73] (as done in [196, Sec. III-A and III-B]) or by considering an average memoryless channel (as done in [135, Sec. III-F], see also [197]). Although both approaches can in principle be used in the context of amplitude-modulated solitons, they both rely on having access to samples from the *true channel*, and not from a (potentially memoryless) model. Such samples can only be obtained through numerical simulations or an optical experiment, which is beyond the scope of this paper. In this context, the channel capacity lower bounds in Sec. 5.6.3, should be considered as a first step towards more involved analyses.

5.6 Non-Gaussian channel law

For the ME case, the conditional PDF for the received soliton amplitude R given the transmitted amplitude A can be obtained by using standard perturbative approach and the Fokker-Planck equation method briefly described in Appendix A. The result can be expressed as a noncentral chi-squared distribution

$$p_{R|A}(r|a) = \frac{1}{\sigma_N^2}\frac{r}{a}\exp\left(-\frac{a+r}{\sigma_N^2}\right) I_2\left(\frac{2\sqrt{ar}}{\sigma_N^2}\right), \tag{5.35}$$

where

$$\sigma_N^2 = D \cdot \frac{Z}{2} \tag{5.36}$$

is the normalised variance of accumulated ASE noise, and $I_2(\cdot)$ denotes the modified Bessel function of the first kind of order two. The expression in (5.35) is nothing else but a special case of a a noncentral chi-squared distribution with six degrees of freedom (see, *e.g.,* [198, Eq. (29.4)]) providing non-Gaussian statistics for Manakov soliton amplitudes. It is convenient to designate the output of the discrete-time channel model as *the square root* of the output soliton amplitudes by making the following change of variables $Y = \sqrt{R}$, and using $X = \sqrt{A}$, the PDF in (5.35) can be expressed as

$$p_{Y|X}(y|x) = \frac{2}{\sigma_N^2}\frac{y^3}{x^2}\exp\left(-\frac{x^2+y^2}{\sigma_N^2}\right) I_2\left(\frac{2xy}{\sigma_N^2}\right), \tag{5.37}$$

which corresponds to the noncentral chi-distribution with six degrees of freedom. An extra factor $2y$ before the exponential function comes from the Jacobian.

For the NLSE case, it is possible to show that the channel law becomes [30, Eq. 8], see also [179, 180]

$$p_{Y|X}(y|x) = \frac{2}{\sigma_N^2}\frac{y^2}{x}\exp\left(-\frac{x^2+y^2}{\sigma_N^2}\right) I_1\left(\frac{2xy}{\sigma_N^2}\right), \tag{5.38}$$

which corresponds to a noncentral chi-distribution with four degrees of freedom.

It should be noted that although we only consider an amplitude modulation A_k (or, in the NFT terms the imaginary part of each discrete eigenvalue), it is possible to include other discrete degrees of freedom corresponding to various soliton parameters in (5.34) in order to improve the achievable information rates. This is, however, beyond the scope of presented research. Furthermore, the channel models presented in this section were obtained by means of a

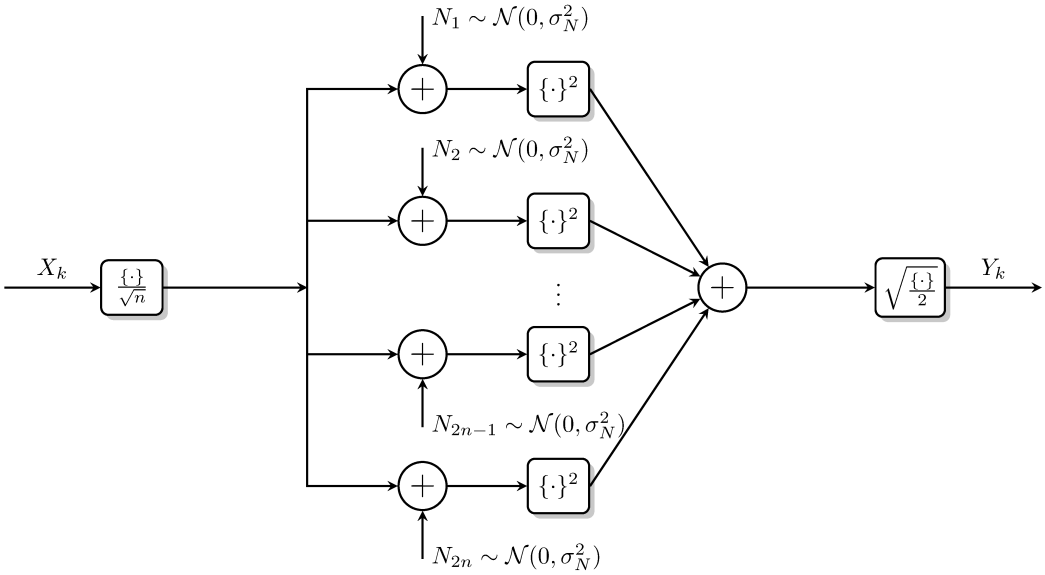

Figure 5.3: Generalised discrete-time channel model: noncentral chi-channel with $2n$ degrees of freedom.

perturbative treatment, and thus, in the context of soliton/eigenvalue communications they are technically valid only at high SNR regime.[9] Despite that, we shall also study capacity lower bounds of a general noncentral chi-channel with arbitrary number of degrees of freedom over any range of SNR. While admittedly the low-SNR region is currently only of interest when $n = 1$ (non-coherent phase channel) we believe its generalisation for $n > 1$ can still be of interest for the new generation of NL optical regeneration systems.

It can be easily observed that both scalar and vector soliton channels can be modelled using the same class of the noncentral chi-distribution with an even number of degrees of freedom $2n$, with $n = 2, 3$. The simplest channel of this type corresponds to $n = 1$, which describes a fibre optical communication channel with zero-dispersion [156] as well as the non-coherent phase channel studied in [199] (see also [200]). Motivated by this, here we consider a general communication channel described by the noncentral chi-distribution with arbitrary (even) degrees of freedom $2n$. Although we are currently not aware of any physically-relevant communication system that can be modelled with $n \geq 4$, we present results for arbitrary n to provide an exhaustive treatment for channels of this type. The channel in question is therefore modelled via

[9] More precisely, when the total soliton energy in the time slot is much greater than that of the ASE noise.

the PDF corresponding to noncentral chi-distribution

$$p_{Y|X}(y|x) = \frac{2}{\sigma_N^2}\frac{y^n}{x^{n-1}}\exp\left(-\frac{x^2+y^2}{\sigma_N^2}\right) I_{n-1}\left(\frac{2xy}{\sigma_N^2}\right), \quad n \in \mathbb{N} \tag{5.39}$$

where $\mathbb{N} \triangleq \{1,2,3,...\}$. This channel law corresponds to the following input-output relation

$$Y^2 = \frac{1}{2}\sum_{i=1}^{2n}\left(\frac{X}{\sqrt{n}} + N_i\right)^2, \tag{5.40}$$

where $\{N_i\}_{i=1}^{2n}$ is a set of independent and identically distributed Gaussian random variables with zero mean and variance σ_N^2. The above input-output relationship is schematically shown in Fig. 5.3, which particularizes to (5.37) and (5.38), for $n = 3$ and $n = 2$, respectively.

5.6.1 Capacity lower bound on scalar channel

In this section, we study capacity lower bounds of the channel law in Eq. (5.38) that describes a scalar memoryless channel. As previously explained, the inter-symbol interference due to pulse interaction can be neglected due to the sufficiently large soliton separation assumed, and thus, the channel can be treated as *memoryless* (see Sec. 5.7 for more details). The channel capacity, in bits per channel use, is then defined by Eq. (2.24), which, it this context, can be re-written as

$$C(\rho) \triangleq \max_{p_X(x):\mathbb{E}[X^2]\leq\sigma_S^2} I_{X,Y}(\rho)\,, \tag{5.41}$$

with the mutual information $I_{X,Y}$ defined by Eq. (2.14). The maximisation in Eq. (5.41) is performed over all possible input statistical distributions $p_X(x)$ that satisfy the power constraint. In our case, this constraint corresponds to a fixed second moment of the input symbol distribution or, equivalently, to a fixed average signal power in a given symbol period. Here we obtain our results as a function of the introduced *effective* SNR defined as $\rho \triangleq \sigma_S^2/\sigma_N^2$, where σ_S^2 is the second moment of the input distribution $p_X(x)$ and σ_N^2 is given by (5.36). The value of σ_S^2 also corresponds to the average soliton amplitude, *i.e.*, $\sigma_S^2 = \mathbb{E}\left[X^2\right] = \mathbb{E}[A]$. It can be shown that for given system parameters, the noise power (in real world units) is constant and proportional to σ_N^2, and the signal power (in real world units) is proportional to σ_S^2. The parameter ρ therefore indeed corresponds to an effective SNR. It is worth mentioning that

the *true* SNR is essentially defined as [135, Eq. (29)]

$$\mathrm{SNR} \triangleq \frac{\mathbb{E}\left[E(A)\right]}{\sigma_N^2 T_s} = 2\kappa \cdot \rho\,, \tag{5.42}$$

where the factor κ denotes the ratio between the available bandwidth and the symbol rate T_s^{-1}. Thus, for a fixed bandwidth and symbol rate, the true SNR is proportional to ρ. We shall henceforth consider the capacity and MI as a function of the effective SNR. The dimensionless energy $E(A_k)$ of the k-soliton waveform in a train can be defined as

$$E(A_k) \triangleq \int_{k-1/2}^{k+1/2} |s_k(t)|^2\, dt\,. \tag{5.43}$$

Specifically, we consider the case when the inter-soliton separation T_s is assumed to be much larger than the typical soliton width (*i.e.,* low duty cycle regime), so the integral in Eq. (5.43) can be taken over $(-\infty, \infty)$. This yields the well-known linear energy-amplitude scaling of the soliton pulse $E(A_k) = 2A_k$. This explains the appearance of the pre-factor of 2 in the SNR definition Eq. (5.42). The minimum inter-soliton separation is then determined by the peak power A_k^2 of each individual soliton, which is, in turn, inversely proportional to the square of its width, *i.e.,* $T_k = A_k^{-1}$.

The exact solution for the power-constrained optimisation problem Eq. (5.41) with the channel law Eq. (5.38) is unknown. In order to obtain a *lower bound* on the capacity C, we shall assume the input symbols X are drawn from a trial input distribution. In this section, we use the Rayleigh PDF

$$p_X(x) = \frac{2x}{\sigma_S^2} \exp\left(-\frac{x^2}{\sigma_S^2}\right), \quad x \in [0, \infty)\,, \tag{5.44}$$

which leads to exponentially-distributed soliton amplitudes A_0 with mean σ_S^2.

The MI in Eq. (5.41) can be decomposed by using Eq. (2.18). The next two Lemmas provide exact closed-form expressions for the output differential entropy h_Y of symbols Y (defined in Eq. (2.8)) with input symbols X distributed according to Eq. (5.44) and for the conditional differential entropy $h_{Y|X}$ (defined in Eq. (2.9)).

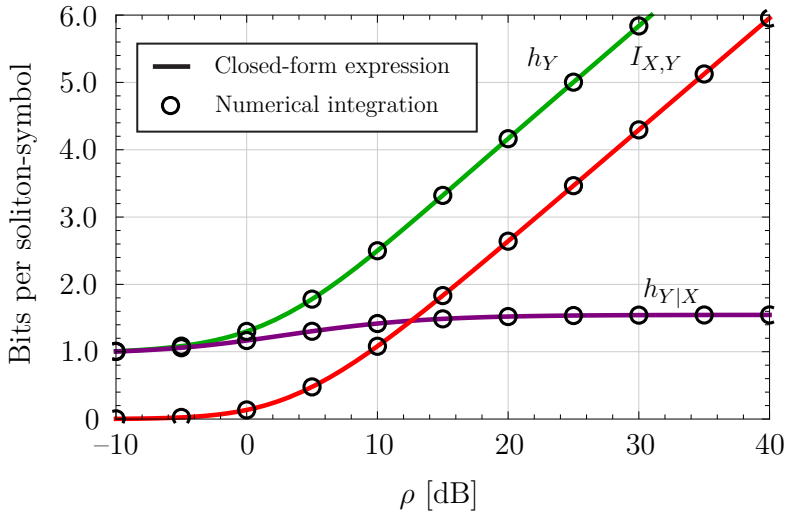

Figure 5.4: The differential entropy of output distribution h_Y in (5.45), the differential conditional entropy $h_{Y|X}$ in (5.46), and the MI I_{XY} in (5.48). Results obtained via numerical integration are also shown (circles).

Lemma 5.6.1. *For the channel in Eq.* (5.38) *and the input distribution Eq.* (5.44)

$$\begin{aligned} h_Y(\rho) = &\log\rho - \left(\frac{1+\rho^{-1}}{2}\right)\log(1+\rho) \\ &+ \psi\left(\rho^{-1}\right) - \frac{3}{2}\psi(1) - \log 2 + 1, \end{aligned} \tag{5.45}$$

where $\psi(x)$ is the digamma function.

Lemma 5.6.2. *For the channel in Eq.* (5.38) *and the input distribution Eq.* (5.44)

$$\begin{aligned} h_{Y|X}(\rho) = &\frac{1}{2}\log\rho - \left(1+\rho^{-1}\right)\log(1+\rho) + 2(\rho+1) \\ &- \rho^{-1}\sqrt{1+\rho^{-1}}\,F_2(\rho) - \frac{\psi(1)}{2} - \log 2, \end{aligned} \tag{5.46}$$

where

$$F_2(\rho) \triangleq \int_0^\infty \xi K_1\left(\sqrt{1+\rho^{-1}}\,\xi\right) I_1(\xi)\log\left[I_1(\xi)\right] d\xi, \tag{5.47}$$

and $I_1(x)$ and $K_1(x)$ are the modified Bessel functions of order one of the first kind and second kind, respectively.

Sketch of the proof. To prove both Lemma 5.6.1 and Lemma 5.6.2, the output distribution $p_Y(y) \triangleq \int_0^\infty p_{Y|X}(y|x)\,p_X(x)\,dx$ is calculated using Eq. (5.38) and

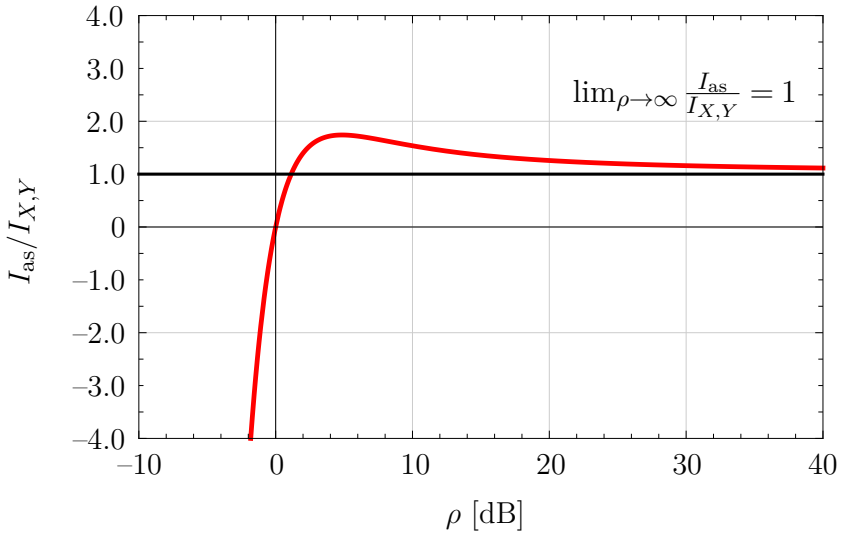

Figure 5.5: The plot of ratio between the MI $I_{X,Y}$ in (5.48) and the function I_{as} in (5.50).

Eq. (5.44). The derived output PDF $p_Y(y)$ is then used in the definitions of differential entropies. The results of both Lemmas are then obtained by evaluating the corresponding integrals. The calculation follows closely that from the earlier work [156], where calculations were performed for a chi-squared distribution with two degrees of freedom (*cf.* (5.47) and [156, Eq. (24)]). The detailed derivations of (5.45) and (5.46) are given in Appendix C. ■

We note that the proof of Lemma 5.6.2 includes finding a closed-form expression for the differential entropy of a chi-squared distribution with four degrees of freedom. To the best of our knowledge, this has never been previously reported in the literature.[10] The results from Lemmas 5.6.1 and 5.6.2 can be combined to produce the following theorem.

Theorem 5.6.3. *For the channel Eq.* (5.38) *and the input distribution Eq.* (5.44)

$$I_{X,Y}(\rho) = \log\left(\rho\sqrt{1+\rho^{-1}}\right) + \frac{\rho^{-1}}{2}\log(1+\rho) - \rho + \rho^{-1}\sqrt{1+\rho^{-1}}\,F_2(\rho) + \psi\left(\rho^{-1}\right) - \psi(1) - 1. \tag{5.48}$$

Proof. From Lemmas 5.6.1 and 5.6.2 and the decomposition in Eq. (2.18). ■

The next theorem shows that the capacity lower bound is asymptotically

[10] However, a closed-form expression for the *expected-log* of a noncentral chi-squared distribution with even number of degrees of freedom was given in [201, Lemma 10.1].

equivalent to half the logarithm of the SNR, which can be considered as the main result.

Theorem 5.6.4. *The MI $I_{X,Y}$ in* (5.48) *satisfies*

$$\lim_{\rho\to\infty} \frac{I_{\text{as}}}{I_{X,Y}(\rho)} = 1, \tag{5.49}$$

where

$$I_{\text{as}} \triangleq \frac{1}{2}\log\rho. \tag{5.50}$$

Proof. The proof follows from an asymptotic expansion of I_{XY} in Eq. (5.48) together with the asymptotic expansion of function Eq. (5.47), *i.e.,* it can be straightforwardly shown that

$$\lim_{\rho\to\infty} \frac{\rho^{-1}\sqrt{1+\rho^{-1}}\,F_2(\rho)}{G_2(\rho)} = 1\,, \tag{5.51}$$

where the asymptotically equivalent function $G_2(\rho)$ is equal to

$$G_2(\rho) = -\frac{1}{2}\log\left(\frac{\rho}{4\pi}\right) + 2\rho - \frac{\psi(1)}{2}\,. \tag{5.52}$$

Then, we can show that

$$\psi\left(\rho^{-1}\right) \sim -\rho\,, \tag{5.53}$$

by proving the following limit

$$\lim_{\rho\to\infty} \frac{\psi\left(\rho^{-1}\right)}{-\rho} = 1\,. \tag{5.54}$$

Accounting that $\lim_{\rho\to\infty}\left[\rho^{-1}\log(1+\rho)\right] = 0$ and $\lim_{\rho\to\infty}\log\left(1+\rho^{-1}\right) = 0$ in combination with Eqs (5.51), (5.53) completes the proof.

■

Fig. 5.5 shows the numerical evaluation of the ratio $I_{\text{as}}/I_{X,Y}(\rho)$ and confirms that the MI behaves asymptotically as $(1/2)\log\rho$, or equivalently, as $(1/2)\log(\text{SNR})$. According to Fig. 5.5, the asymptotic function (5.50) approaches the MI from above. Interestingly, the expression (5.50) has appeared in asymptotic analyses of optical systems (see, *e.g.,* [156, eq. (25)], [170, Sec. V-A], [202, eq. (6)]). Since the channel capacity is lower-bounded by $I_{X,Y}(\rho)$, this result implies that the capacity grows at least as fast as $(1/2)\log(\text{SNR})$,

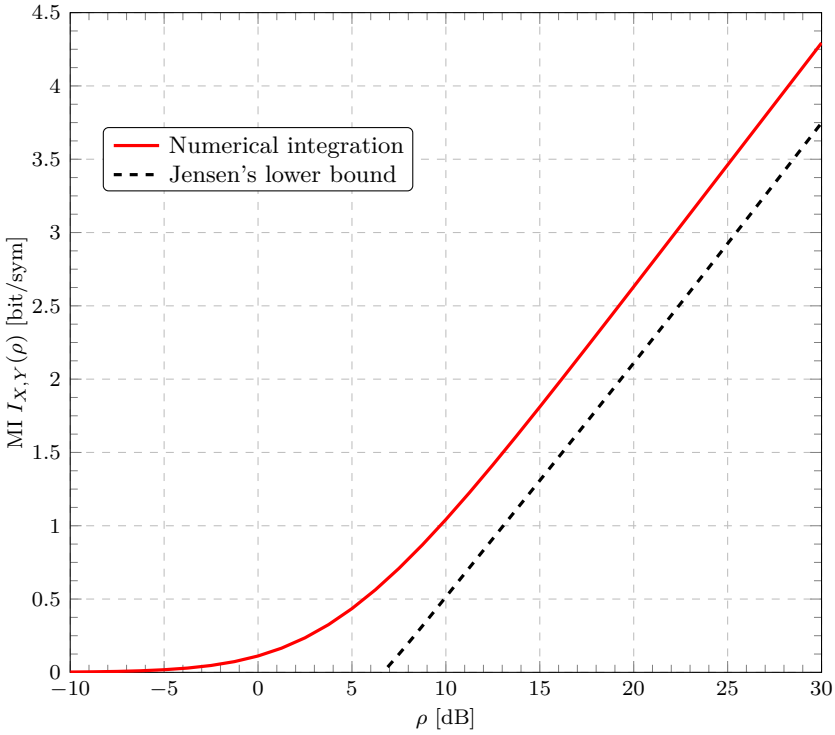

Figure 5.6: Analytical lower bound on the MI based on Jensen's inequality, *i.e.*, $I_{X,Y}(\rho) = h_Y^{(\text{Jensen})} - h_{Y|X}$ given by the entropies in Eqs (5.55) and (5.57).

when SNR $\to \infty$.

5.6.2 Jensen's lower bound on vector Manakov channel

In this short section, we demonstrate the possibility to analytically lower-bound the MI for the Manakov channel given by Eq. (5.4). Unfortunately, the MI does not have an exact analytical closed form solution. The next theorem, however, demonstrates the non-decaying lower bound on the MI, which is derived by utilising the well-known Jensen's inequality [32].

Lemma 5.6.5. *For the channel in Eq.* (5.37)*, the lower bound on output entropy* h_Y *reads*

$$h_Y \geq h_Y^{(\text{Jensen})} = -\log\left\{\frac{2(\rho+1)\,\mu(\rho)}{\rho^2}\left[e^{-\frac{\mu^2(\rho)}{\rho+1}} - e^{-\mu^2(\rho)}\right] - \frac{2\mu^3(\rho)}{\rho}e^{-\mu^2(\rho)}\right\}, \tag{5.55}$$

where

$$\mu(\rho) = \frac{\sqrt{\pi}}{2\rho^2}\left[(\rho+1)^{5/2} - (5/2)\rho - 1\right]. \tag{5.56}$$

The next theorem indicates that, in the case of Manakov channel, the conditional differential entropy has exact closed form solution.

Lemma 5.6.6. *For the channel in Eq.* (5.37)*, the conditional entropy* $h_{Y|X}$ *is equal to*

$$h_{Y|X} = -\frac{3}{2}\log\left(1+\rho^{-1}\right) - \frac{3}{2}\left(\frac{2\rho+1}{\rho^2}\right)\log(1+\rho) + 2\rho + \frac{3\rho^{-1}}{2} - \rho^{-1}\left(1+\rho^{-1}\right)F_2(\rho) - \frac{\psi(1)}{2} - \log 2 + 3\,. \tag{5.57}$$

Proof. The detailed derivations of the lower bound in (5.55) and closed form solution of (5.57) are given in Appendix D. ■

Fig. 5.6 illustrates the resulting MI given by the entropies in Eqs. (5.55) and (5.55). Using asymptotic expansions, similar to the proof of Theorem 5.6.4, it can be straightforwardly shown that the Jensen's lower bound on the mutual information grows unbounded as SNR tends to infinity, *i.e.,*

$$I_{X,Y} \geq I_{X,Y}^{(\text{Jensen})} = \frac{1}{2}\log\rho + O[1] \quad \rho \to \infty\,. \tag{5.58}$$

It can be observed that for $\rho > 10$ dB, there is a constant gap between the analytical lower bound and the exact numerical Monte Carlo integration, which is approximately equal to 0.52 bit per soliton symbol. Therefore, despite the advantages of analytical closed form, the monotonically increasing Jensen's lower bound cannot be considered as a tight bound. In the next section, we show the asymptotically tight analytical lower bound for the generalised channel model.

5.6.3 Bounds of generalised channel

In this section, we develop an analytical lower bound on the capacity on the generalised channel law given by Eq. (5.39) . We do not aim at finding the capacity-achieving distribution, but instead, we study lower bounds on the capacity. We do this because the capacity problem is in general very difficult, but also because of the relevance of having non-decreasing lower bounds on the capacity for the optical community. We emphasise once again that to obtain a lower bound on the capacity, we will simply choose an input distribution $p_X(x)$ (as done in, *e.g.,* [90, 30]). Without claiming the generality, we, however, consider four important candidates for the input distribution. As we will see later, this Rayleigh input distribution Eq. (5.44) is not the one giving the highest lower bound. However, it has one important advantage: it allows some

analytical results for the mutual information. The other three distributions are considered later in this section as numerical examples.

The next two Lemmas provide an exact closed-form expression for the conditional differential entropy $h_{Y|X}(\rho)$ and an asymptotic expression for the output differential entropy $h_Y(\rho)$.

Lemma 5.6.7. *For the channel in Eq.* (5.39) *and the input distribution Eq.* (5.44)

$$\begin{aligned} h_{Y|X}(\rho) = &\left(2\rho + n - \frac{n}{2}\psi(n)\right)\log_2 e - 1 + \frac{n-1}{2}\left(\log_2\rho + \psi(1)\log_2 e\right) \\ &- \frac{n\log_2 e}{2}\frac{\rho}{\rho+1}\,\Phi\left(\frac{\rho}{\rho+1}, 1, n\right) \\ &- \rho^{-1}\left(\frac{\rho+1}{\rho}\right)^{(n-1)/2} F_n(\rho)\log_2 e\,, \end{aligned} \tag{5.59}$$

where $\psi(x) \triangleq d\log\Gamma(x)/dx$ *is the digamma function and* $\Phi(\alpha, 1, n)$ *is the special case of the Lerch transcendent function [117, Eq. (9.551)]*

$$\Phi(\alpha, 1, n) \triangleq -\frac{\log(1-\alpha)}{\alpha^n} - \sum_{k=0}^{n-2}\frac{\alpha^{k+1-n}}{k+1}\,. \tag{5.60}$$

The function $F_n(\rho)$ *is defined as*

$$F_n(\rho) \triangleq \int_0^\infty \xi K_{n-1}\left(\sqrt{1+\rho^{-1}}\,\xi\right) I_{n-1}(\xi)\log\left[I_{n-1}(\xi)\right] d\xi\,, \tag{5.61}$$

and $K_n(x)$ *is the modified Bessel function of the second kind of order* n*.*

Proof. See Appendix E ■

Lemma 5.6.8. *For the channel in* (5.39) *and the input distribution* (5.44)

$$h_Y(\rho) = \frac{1}{2}\log_2\rho + \left(1 - \frac{\psi(1)}{2}\right)\log_2 e - 1 + O\left[\rho^{-1}\right], \quad \rho \to \infty \tag{5.62}$$

Proof. See Appendix E ■

The next theorem is one of the main results of this chapter.

Theorem 5.6.9. *The MI for the channel in* (5.39) *and the input distribution*

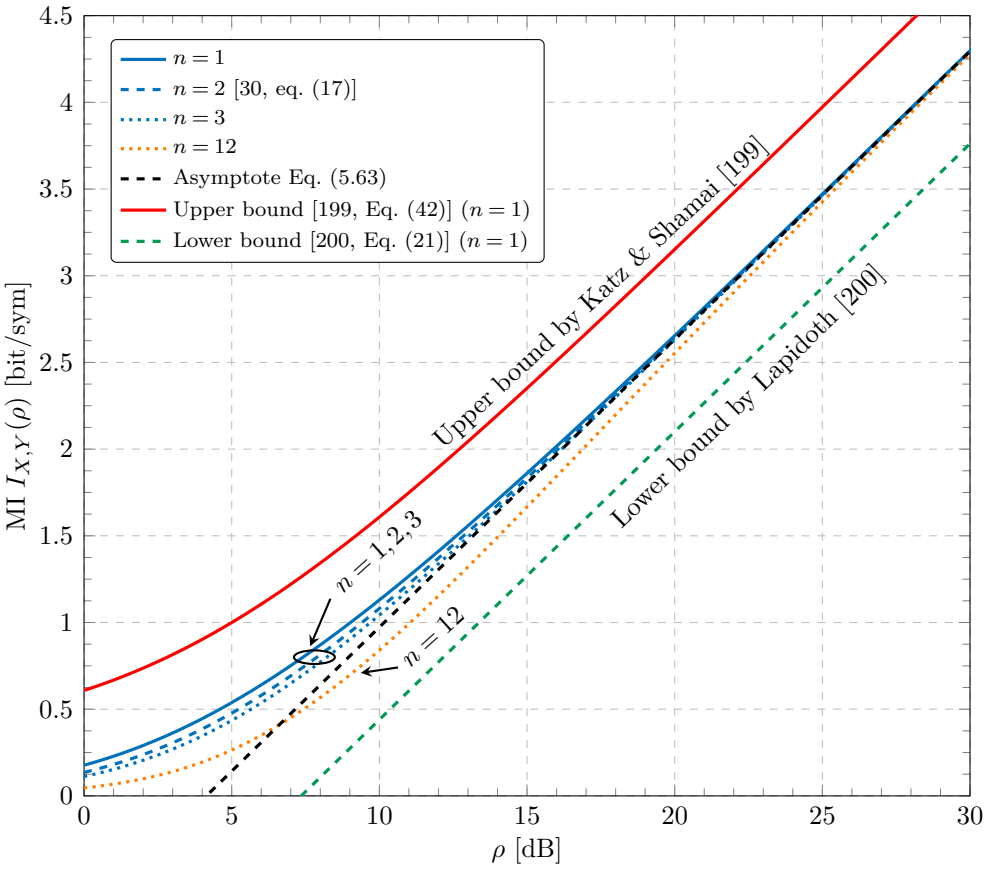

Figure 5.7: The MI $I_{X,Y}(\rho)$ in (2.14) (numerically calculated) for the chi-distribution with different degrees of freedom and the channel model (5.39). The asymptotic estimate given by Theorem 5.6.9 is also shown. Lower and upper bounds for $n=1$ are also shown.

(5.44) *admits the following asymptotic expansion*

$$I_{X,Y}(\rho) = \frac{1}{2}\log_2\left(\frac{e^{1-\psi(1)}}{4\pi}\rho\right) + O\left[\rho^{-1}\right], \quad \rho\to\infty. \tag{5.63}$$

Proof. We expand the function $F_n(\rho)$ in (5.47) defining the conditional entropy in Lemma 5.6.7. At fixed large ρ the integrand asymptotically decays as $\exp(-\xi/2\rho)$, i.e., with small decrement (which can be proven by standard large argument asymptotes of the Bessel functions). This means that the main contribution to the integral comes from the asymptotic region $1 \lesssim \xi \lesssim \rho$ in most part of which the large argument expansion of both Bessel functions is indeed justified. Using it uniformly we obtain

$$F_n(\rho) = 2\rho^2 + \frac{\rho}{2}\left[\log\frac{1}{\rho} + 1 - \log 4\pi - \psi(1)\right] + O[1],$$

which used in (5.59) gives the asymptotic expression

$$h_{Y|X}(\rho) = \frac{1}{2}\log_2 \pi e + O\left[\rho^{-1}\right], \quad \rho \to \infty. \tag{5.64}$$

The proof is completed by combining (5.64) and (5.62) with (2.14). ■

The result in Theorem 5.6.9 is a universal and n-independent expression. The expression in Eq. (5.63) shows that the capacity lower bound is asymptotically equivalent to half of logarithm of SNR plus a constant which is order-independent. Fig. 5.7 shows the numerical evaluation of $I_{X,Y}(\rho)$ for $n = 1, 2, 3, 12$ obtained by numerically evaluating all the integrals in the exact expressions for the conditional and output entropies in Eq. (5.59) and Eq. (E.18), as well as the asymptotic expression in Theorem 5.6.9. Interestingly, we can see that even in the medium-SNR region, the influence of the number of degrees of freedom on the MI is minimal, and the curves are quite close to each other. In this figure, we also include the lower and upper bounds for $n = 1$ given by [200, Eq. (21)] and [199, Eq. (41)], respectively. These results show that the asymptotic results in Theorem 5.6.9 correctly follow these two bounds.

Note that, the main reason for considering a Rayleigh input distribution was that it yields a semi-analytical lower bound on the the capacity. In the following example, we consider three other input distributions and numerically calculate the resulting MI.

Example 1. *Consider the geometric (exponential), half-Gaussian, and Maxwell-Boltzmann distributions given by*

$$p_X(x) = \frac{\sqrt{2}}{\sigma_S}\exp\left(-\frac{\sqrt{2}x}{\sigma_S}\right), \quad x \in [0, \infty), \tag{5.65}$$

$$p_X(x) = \frac{\sqrt{2}}{\sqrt{\pi}\sigma_S}\exp\left(-\frac{x^2}{2\sigma_S^2}\right), \quad x \in [0, \infty), \tag{5.66}$$

and

$$p_X(x) = \frac{3\sqrt{6}x^2}{\sqrt{\pi}\sigma_S^3}\exp\left(-\frac{3x^2}{2\sigma_S^2}\right), \quad x \in [0, \infty), \tag{5.67}$$

respectively.

The MIs for these three distributions for $n = 1, 2, 3$ are shown in Fig. 5.8 and show that the lower bound given by the geometric input distribution in (5.65) displays high MI in the low SNR regime ($\rho < 10$ dB), whereas the half-Gaussian input distribution in (5.66) is better for medium and large SNR.

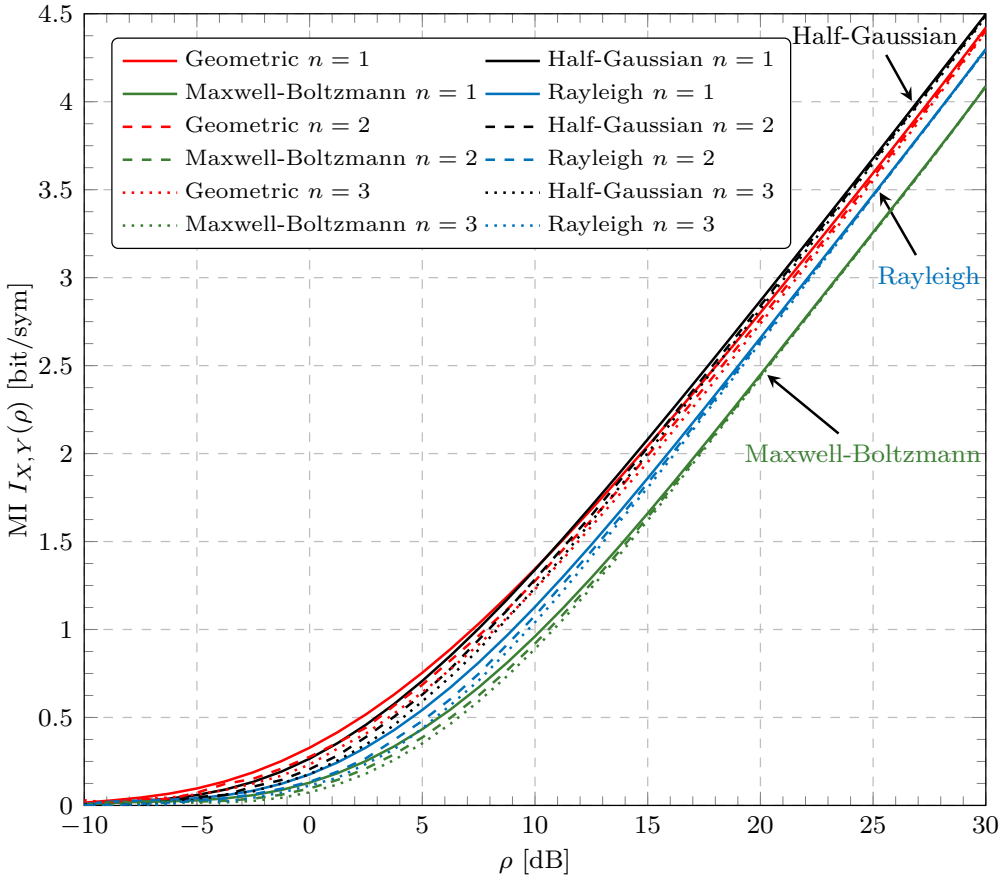

Figure 5.8: MI estimates (by numerically evaluating (2.14) via Monte-Carlo integration) for different trial continuous input distributions and different values of n (different line types). Different distributions are shown with different colours.

On the other hand, the Maxwell-Boltzmann distribution in (5.67) gives the lowest MI for all SNR. Numerical results also indicate that all the presented MIs asymptotically exhibit an equivalent growth irrespective of the number of the degrees of freedom $2n$. The following example considers the use of discrete constellations. In particular, we assume that the soliton amplitudes take values on a set $\mathcal{X} \triangleq \{x_1, ..., x_M\}$, where $M \triangleq |\mathcal{X}| = 2^m$ is the cardinality of the constellation, and m is a number of bits per symbol. The MI (2.14) in this case can be evaluated as

$$I_{X,Y}(\rho) = \frac{1}{M} \sum_{x \in \mathcal{X}} \int_0^\infty p_{Y|X}(y|x) \log_2 \frac{p_{Y|X}(y|x)}{\frac{1}{M} \sum_{x' \in \mathcal{X}} p_{Y|X}(y|x')} \, dy, \tag{5.68}$$

where we assumed the symbols are equally likely.

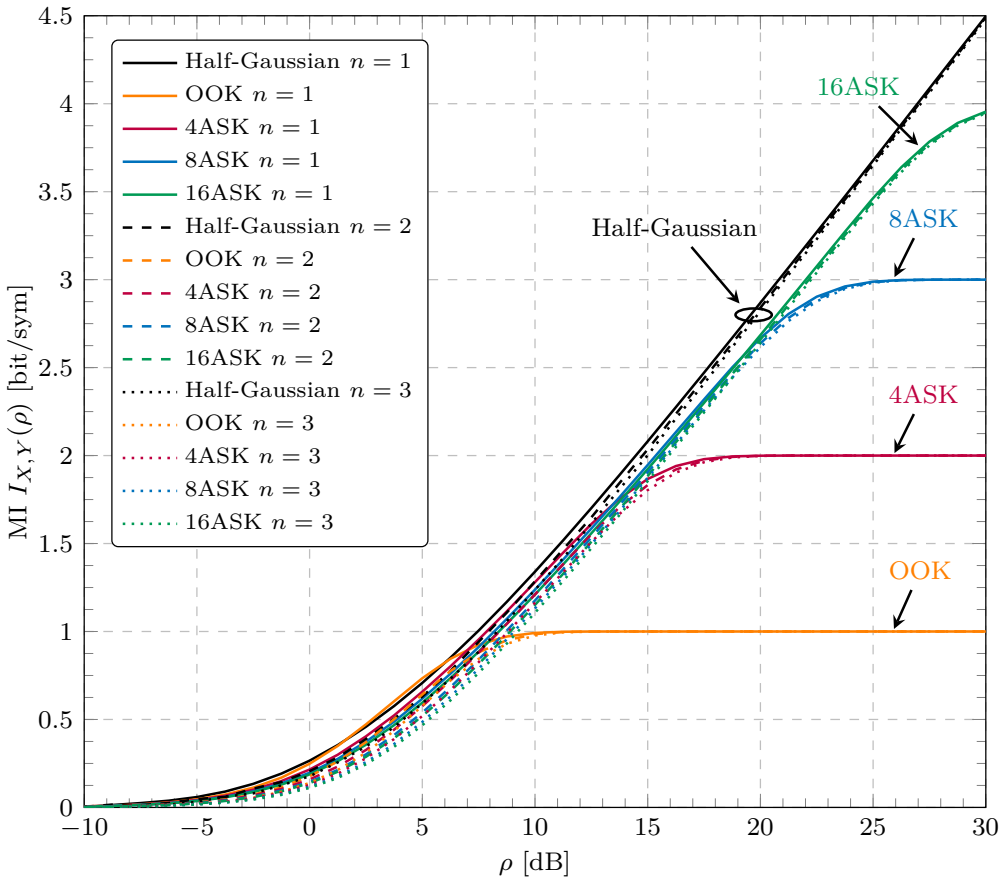

Figure 5.9: MI estimates (numerically calculated) for equally-spaced M-ASK signal constellations with $M = \{2,4,8,16\}$ constellation points.

Example 2. *Let us consider amplitude-shift keying (ASK) constellations* $\mathcal{X} = \{0,1,\ldots,M-1\}$ *with* $m = 1,2,3,4$ *and second moment* σ_S^2*, which correspond to OOK, 4-ASK, 8-ASK, and 16-ASK, respectively.*

The MI numerically evaluated for these constellations is shown in Fig. 5.9 for chi-channel with $n = 1,2,3$. As a reference, in this figure we also show (black lines) the MI for the (continuous) half-Gaussian input distribution. The results in this figure show that in the low SNR regime, the use of binary modulation is in fact better than the half-Gaussian distribution. This can, however, be remedied by using a geometric distribution, which, as shown in Fig. 5.8, outperforms the half-Gaussian distribution in the low SNR regime. In the high SNR regime, however, this is not the case.

Table 5.1: Simulation system parameter

Parameter	Value
Carrier frequency (ν_0)	193.41 THz
Fibre attenuation (α_{fibre})	$0.20\,\text{dB km}^{-1}$
Fibre group-velocity dispersion (β_2)	$-21.67\,\text{ps}^2\,\text{km}^{-1}$
Fibre nonlinearity (γ)	$2.0\,\text{W}^{-1}\text{km}^{-1}$
Phonon occupancy factor (K_T)	1.13
Propagation distance	500, 2000 km
Propagation step-size	0.1 km

5.7 Memoryless property of the discrete-time channel model

In this section, we present numerical simulations to verify the memoryless assumption for the discrete channel model in Sec. 5.5. To this end, we simulated the propagation of sequences of, for instance, $N = 10$ soliton symbols through the scalar waveform channel given by Eq. (5.5). Two launched powers (-1.5 and 1.45 dBm) and two propagation distances (500 km and 2000 km) are considered. The simulations were carried out via the standard SSFT method. The soliton amplitudes were generated as IID samples from a Rayleigh input distribution (see Eq. (5.44)) and the variance of X was chosen to be 1.25 and 20, so that the resulting soliton waveforms have powers of -1.5 and 1.45 dBm, respectively. The transmitted waveform $x(\tau)$ was created using (5.33) at a symbol rate of 1.7 GBd. In order to guarantee an accurate simulation, the time-domain samples were taken every 4.6 ps and the step size was 0.1 km. White Gaussian noise was added at each step to model the IDRA process. The simulation parameters are similar to those used in [173] and are summarised in Table 5.1.

Fig. 5.10 shows the waveforms before and after propagation through the channel given in (5.5). As expected, the received signal is a noisy version of the transmitted waveform, where the noise increases as the propagation distance increases. These results show that doubling the transmission distance and/or (approximately) doubling the launch power has very little effect in the soliton shapes. The noisy waveforms shown in Fig. 5.10 were then used to obtain soliton amplitudes $\mathbf{Y} \triangleq [Y_1, Y_2, \ldots, Y_{10}]$ via the forward NFT. Each amplitude is obtained by processing the corresponding symbol period via the spectral matrix method [203, Sec. IV-B]. Particularly, the numerical computations of the discrete eigenvalues of the Zakharov-Shabat system of equations Eq. (5.11)

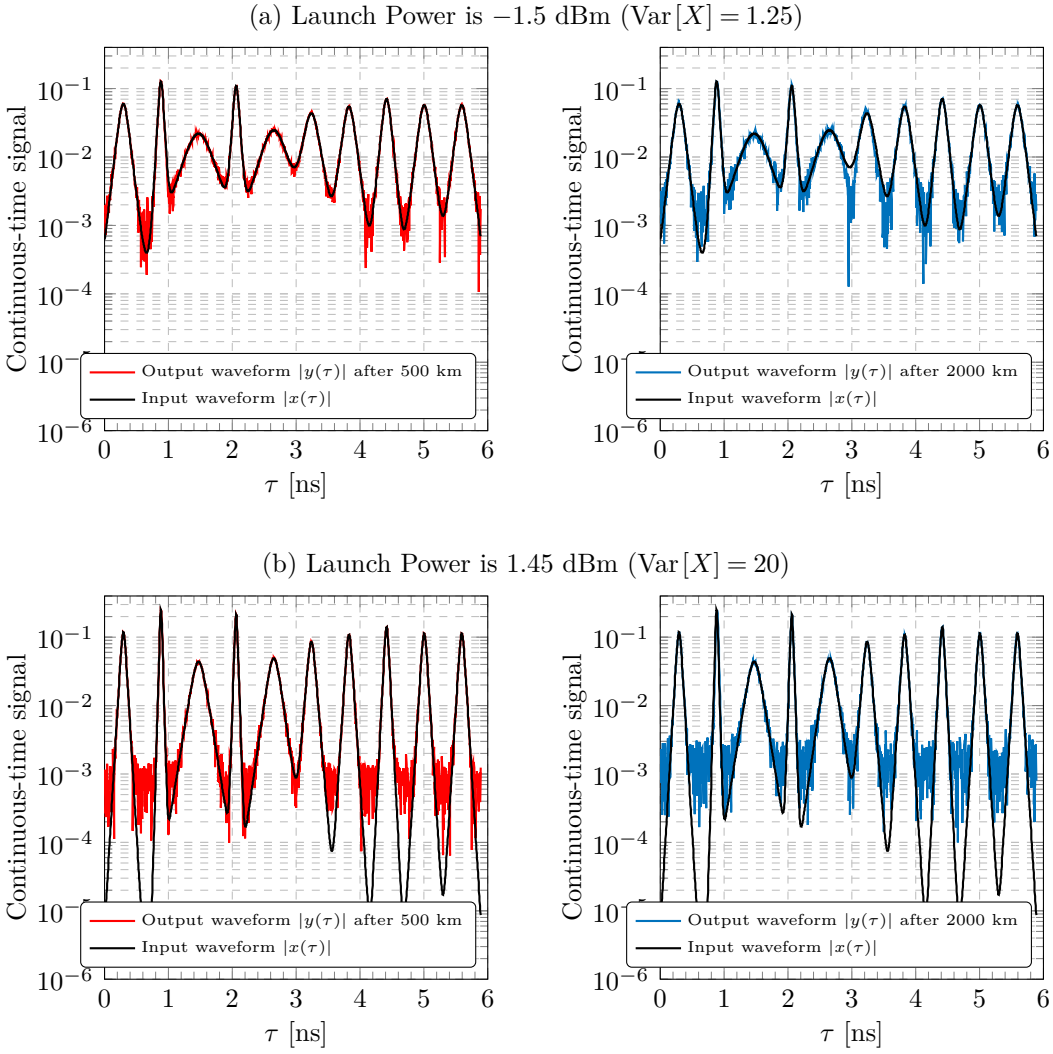

Figure 5.10: Continuous-time input $x(\tau)$ and output $y(\tau)$ soliton waveforms for 10 solitons and distributed noise due to DRA. Two launch powers are considered: (a) −1.5 dBm and (b) 1.45 dBm. The solitons are propagated 500 and 2000 km.

have been done by means of the so-called *Fourier collocation method* [204] (see also Appendix B). A brief description of this quite accurate and reliable method is given in Appendix B. In order to test the memoryless assumption, we perform a simple correlation test. In particular, we consider the normalised output symbol correlation matrix, whose entries are defined as

$$c_{kk'} \triangleq \frac{\mathbb{E}\left[\left(Y_k - \mathbb{E}\left[Y_k\right]\right)\left(Y_{k'} - \mathbb{E}\left[Y_{k'}\right]\right)\right]}{\mathbb{E}\left[Y_k\right]\mathbb{E}\left[Y_{k'}\right]}. \tag{5.69}$$

The obtained correlation matrices are shown in Fig. 5.11, where statistics were gathered by performing 10^3 Monte-Carlo runs of the signal propagation. As we can see from Fig. 5.11, the matrices are almost diagonal. Since our communication channel is believed to be non-Gaussian, the absence of correlation does not, of course, necessarily imply the memoryless property (understood here as the statistical independence). However, it does constitute an important quantification of the qualitative criterion $\exp(-A_k T_s) \ll 1$ as given in Sec. 5.5.

5.8 Asymptotically vanishing rate loss

Let us now address the impact of the cut-off $\hat{a}$ we have introduced in Sec. 5.5. All our results for continuous input distributions have been obtained for the input distributions that are not bounded away from zero (see Eqs (5.44), (5.65), (5.66), (5.67)). Therefore, symbols X_k are generated below the threshold $\hat{x} = \sqrt{\hat{a}}$, where the channel law we considered does not hold. We shall now only examine here the case of the Rayleigh input Eq. (5.44) as this distribution was used to obtain the main result of this chapter. We will prove that in the high-power (*i.e.,* high SNR) regime, the effect of the cut-off on the achievable data rate tends to zero. To do so, we note that for fixed fibre parameters and propagation distance, the cut-off $\hat{x}^2 = \hat{a} = \max[\hat{a}_{\text{noise}}, \hat{a}_{\text{inter}}]$ is also supposed to be fixed, while $\sigma_S^2 = \rho\sigma_N^2$ grows linearly with SNR. In other words, one can achieve high SNR at the expense of high power solitons for fixed noise variance. One possible way of showing that the effect of the cut-off on the achievable rate is zero as SNR tends to infinity is to consider a transmitter which generates a dummy symbol every time $X_k \leq \hat{x}$. The value of the threshold $\hat{x}$ is message-independent, and thus, can be assumed to be known to the receiver which will discard sub-threshold symbols. This allows us to keep the main results of the paper at the expense of a data rate loss (since part of the time, dummy symbols are transmitted). The probability of such "outage" event η is given by an the integral of the input distribution from zero to the threshold. For the Rayleigh input PDF Eq. (5.44) this probability is given by

$$\eta = 1 - \exp\left(-\frac{\hat{a}}{\sigma_S^2}\right). \tag{5.70}$$

Therefore, asymptotically $\eta(\rho) \approx \hat{a}/\left(\rho\sigma_N^2\right) \to 0$ when $\rho \to \infty$. The average rate loss is then given by $1 - \eta(\rho)$, which obviously tends to zero as $\rho \to \infty$.

An alternative and more mathematically rigorous solution to the prob-

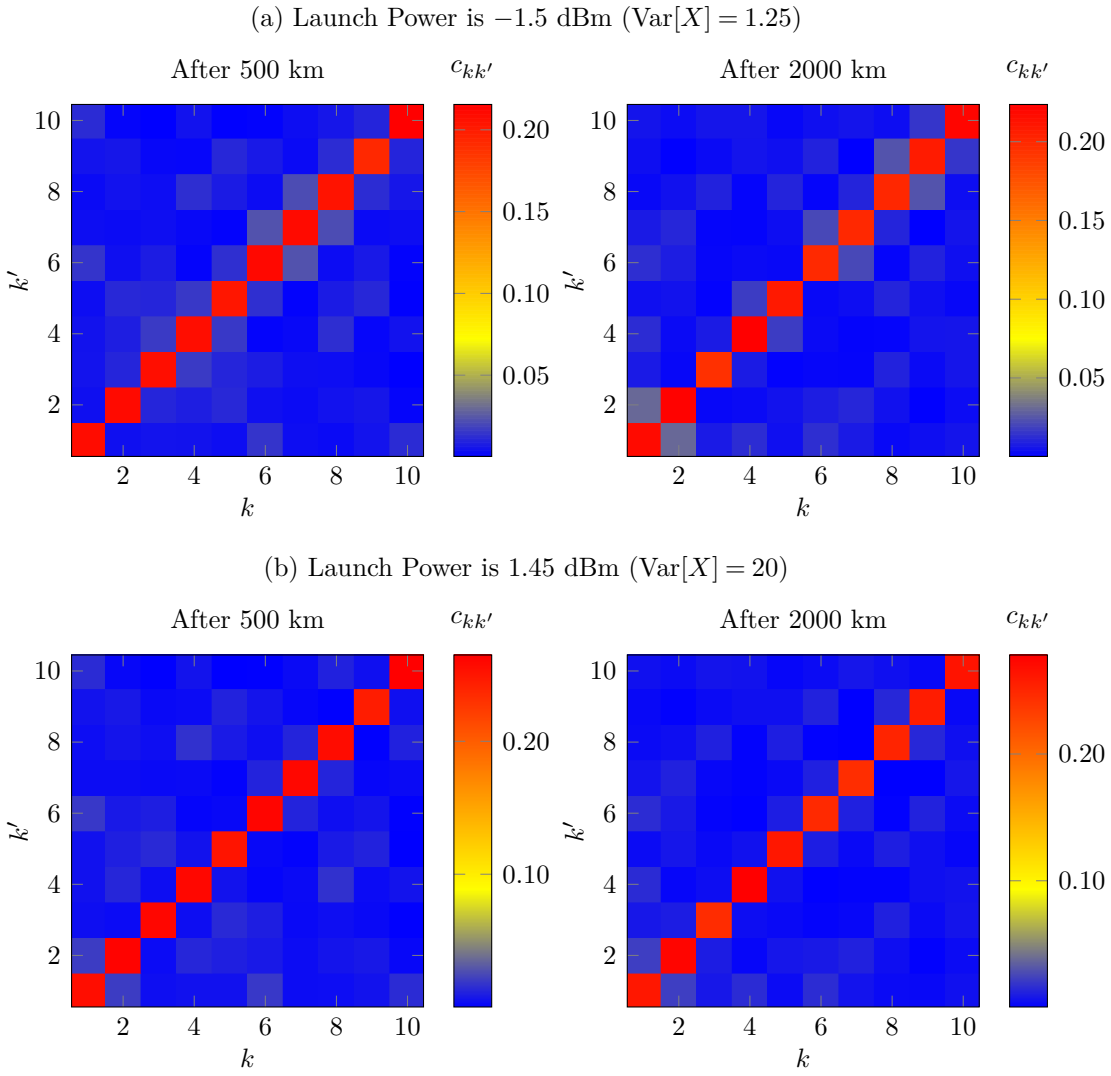

Figure 5.11: Normalised output symbol correlation matrices for the two launch powers and propagation distances in Fig. 5.10.

lem mentioned above is to consider directly the difference between the MI asymptote obtained in the current chapter (see *i.e.*, Theorem 5.6.9), and that obtained by a truncated input Rayleigh distribution which simply does not generate sub-threshold symbols. This difference can be shown to tend to zero as $\rho \to \infty$.

Here, we shall prove that considering a more realistic signalling scheme with an input distribution bounded (truncated) away from zero gives the same results as Theorem 5.6.9 in the limit of large average power $\sigma_S \to \infty$. To this end, consider a system where the transmitted amplitudes X are drawn from a Rayleigh distribution with PDF given in (5.44). Let us now introduce a

threshold $\hat{x}$ of amplitudes realisations below which our channel law model is expected to be inapplicable. Let us now introduce an alternative system where the symbols $\tilde{X}$ are drawn from a "truncated" Rayleigh distribution with PDF

$$p_{\tilde{X}}(x) = \frac{1}{1-\eta} p_X(x)\, H(x-\hat{x}), \qquad x \in [\hat{x}, \infty), \tag{5.71}$$

where $H(x-\hat{x})$ is the Heaviside step function, and η is defined as

$$\eta \triangleq \mathbb{P}[X < \hat{x}]. \tag{5.72}$$

This probability can be expressed as

$$\eta = \int_0^{\hat{x}} p_X(x)\, dx = \frac{2}{\sigma_S^2} \int_0^{\hat{x}} x \exp\left(-\frac{x^2}{\sigma_S^2}\right) dx \tag{5.73}$$

$$= 1 - \exp\left(-\frac{\hat{x}^2}{\sigma_S^2}\right). \tag{5.74}$$

As discussed in Sec. 5.5, the threshold $\hat{x}$ is a constant, and thus, $\lim_{\sigma_S \to \infty} \eta = 0$.

In order to prove that the rate loss tends to zero, we shall prove that $\lim_{\sigma_s \to \infty} \left[I_{X,Y} - I_{\tilde{X},\tilde{Y}} \right] = 0$, or equivalently,

$$\lim_{\sigma_s \to \infty} \left[h_Y - h_{\tilde{Y}} \right] = 0 \tag{5.75}$$

and

$$\lim_{\sigma_s \to \infty} \left[h_{\tilde{Y}|\tilde{X}} - h_{Y|X} \right] = 0. \tag{5.76}$$

To prove (5.75), we have the following:

$$p_{\tilde{Y}}(y) = \int_0^{\infty} p_{Y|X}(y|x)\, p_{\tilde{X}}(x) dx \tag{5.77}$$

$$= \frac{1}{1-\eta} \int_{\hat{x}}^{\infty} p_{Y|X}(y|x)\, p_X(x)\, dx \tag{5.78}$$

$$\leq \frac{1}{1-\eta} \int_0^{\infty} p_{Y|X}(y|x)\, p_X(x)\, dx \tag{5.79}$$

$$= \frac{1}{1-\eta} p_Y(y). \tag{5.80}$$

The Kullback-Leibler divergence (relative entropy) between the distributions $p_{\tilde{Y}}(y)$ and $p_Y(y)$ is defined as

$$D\left(p_{\tilde{Y}}(y)\,\|\,p_Y(y)\right) \triangleq \mathbb{E}\left[\log\frac{p_{\tilde{Y}}(Y)}{p_Y(Y)}\right] \tag{5.81}$$

$$= \int_0^\infty p_{\tilde{Y}}(y)\log\frac{p_{\tilde{Y}}(y)}{p_Y(y)}\,dy \tag{5.82}$$

$$\leq \log\frac{1}{1-\eta}\int_0^\infty p_{\tilde{Y}}(y)\,dy \tag{5.83}$$

$$= -\log(1-\eta) \tag{5.84}$$

$$= \frac{\hat{x}^2}{\sigma_S^2}. \tag{5.85}$$

Using the nonnegativity property of the relative entropy together with (5.85), we obtain

$$\lim_{\sigma_S\to\infty} D\left(p_{\tilde{Y}}(y)\,\|\,p_Y(y)\right) = 0\,. \tag{5.86}$$

Using the fact that the relative entropy is zero if and only if $p_{\tilde{Y}}(y) = p_Y(y)$ *almost everywhere* [32, Theorem 8.6.1], we conclude that (5.75) is fulfilled since the integrands in the differential entropy integrals differ on a set with measure zero.

Let us now turn to the first conditional differential entropy in (5.76), for which we have

$$h_{\tilde{Y}|\tilde{X}} \triangleq -\int_0^\infty\int_0^\infty p_{Y|X}(y|x)\,p_{\tilde{X}}(x)\log p_{Y|X}(y|x)\,dxdy \tag{5.87}$$

$$= \int_0^\infty p_{\tilde{X}}(x)\,g(x)\,dx\,, \tag{5.88}$$

where

$$g(x) \triangleq -\int_0^\infty p_{Y|X}(y|x)\log p_{Y|X}(y|x)\,dy \tag{5.89}$$

represents the conditional differential entropy of $p_{Y|X}(y|x)$, and $p_{Y|X}(y|x)$ is given by the noncentral chi-distribution (5.39).

Using (5.71), the conditional differential entropy $h_{\tilde{Y}|\tilde{X}}$ can be expressed as

$$h_{\tilde{Y}|\tilde{X}} = \frac{1}{1-\eta} \int_{\hat{x}}^{\infty} p_X(x)\, g(x)\, dx, \tag{5.90}$$

$$= \frac{1}{1-\eta} \left[\int_{0}^{\infty} g(x)\, p_X(x)\, dx - \int_{0}^{\hat{x}} g(x)\, p_X(x)\, dx \right] \tag{5.91}$$

$$= \frac{h_{Y|X}}{1-\eta} - \frac{1}{1-\eta} \int_{0}^{\hat{x}} g(x)\, p_X(x)\, dx\,. \tag{5.92}$$

The first term on the right-hand side of (5.92) tends to the conditional entropy of the untruncated distribution. We shall now prove that the last (integral) term in (5.92) tends to zero when $\sigma_S \to \infty$. We note that according to (5.44) the input distribution $p_X(x)$ tends to zero uniformly in the interval $[0, \hat{x}]$ as $\sigma_S \to \infty$. Then, according to the bounded convergence theorem, in order to prove that integral term in (5.92) is asymptotically vanishing, it is sufficient to prove that the function $g(x)$ remains bounded within the interval $[0, \hat{x}]$. We shall do so by providing separate upper and lower bounds for this function.

The upper bound for $g(x)$ can be obtained by considering a relative entropy between the channel law $p_{Y|X}(y|x)$ and an auxiliary distribution $p_Y^{\diamond}(y)$ supported on $[0, \infty)$. Since the relative entropy is always positive (see Eq. (2.12)), it readily provides an upper bound for the differential entropy (5.89), namely,

$$g(x) \leq -\mathbb{E}\left[\log p_Y^{\diamond}(Y)\right] = -\int_{0}^{\infty} dy\, p_{Y|X}(y|x) \log p_Y^{\diamond}(y)\,. \tag{5.93}$$

Choosing a half-Gaussian distribution

$$p_Y^{\diamond}(y) = \frac{2}{\sqrt{\pi}} \exp\left(-y^2\right),$$

immediately gives an upper bound

$$g(x) \leq \mathbb{E}\left[Y^2\right] - \log\left(\frac{2}{\sqrt{\pi}}\right).$$

The second-moment for the noncentral chi distribution is readily available, *e.g.,* from (5.40), leading to the following upper bound:

$$g(x) \leq x^2 + n\,\sigma_N^2 + \log\frac{\sqrt{\pi}}{2}. \tag{5.94}$$

Note that this upper bound is bounded inside an arbitrary finite interval $[0, \hat{x}]$.

Establishing a lower bound for $g(x)$ is slightly more involved. The first step is to transform the noncentral chi distribution into a noncentral chi-squared distribution by making the following change of variable in the integral (5.89): $z = 2y^2/\sigma_N^2$. Introducing the additional notation $\lambda = 2x^2/\sigma_N^2$ and $n = k/2$, where k is a number of degrees of freedom of noncentral chi-squared distribution, we obtain

$$p_{Z|\Lambda}(z|\lambda) = \frac{1}{2}\left(\frac{z}{\lambda}\right)^{(k-2)/4} \exp\left(-\frac{z+\lambda}{2}\right) I_{(k-2)/2}\left(\sqrt{\lambda z}\right), \quad z \in [0,\infty). \tag{5.95}$$

We can now express $g(x)$ in (5.89) as an average with respect to the noncentral chi-squared distribution:

$$g(\lambda) = -\int_0^\infty p_{Z|\Lambda}(z|\lambda) \log\left[\frac{2^{3/2}\,z^{1/2}}{\sigma_N}\, p_{Z|\Lambda}(z|\lambda)\right] dz \tag{5.96}$$

$$= g^{(1)}(\lambda) + g^{(2)}(\lambda) + \frac{3}{2}\log 2 - \log\sigma_N\,, \tag{5.97}$$

where we have introduced two functions: $g^{(1)}(\lambda)$, which represents the differential entropy of the noncentral chi-squared distribution $p_{Z|\Lambda}(z|\lambda)$, i.e.,

$$g^{(1)}(\lambda) \triangleq -\int_0^\infty dz\, p_{Z|\Lambda}(z|\lambda)\, \log p_{Z|\Lambda}(z|\lambda)\,, \tag{5.98}$$

and $g^{(2)}(\lambda)$, which stands for minus half of the so-called *expected-log*, *i.e.,*

$$g^{(2)}(\lambda) \triangleq -\frac{1}{2}\,\mathbb{E}\,[\log Z\,]. \tag{5.99}$$

The motivation for the above transformation stems from the fact that it has been proven in [205] that the noncentral chi-squared distribution function (5.95) is *log-concave* (*i.e.,* log of $p_{Z|\Lambda}(z|\lambda)$ is concave) if the number of degrees of freedom $k \geq 2$, *i.e.,* $n \geq 1$, which is always the case. On the other hand, the differential entropy of any log-concave distribution function can be lower-

bounded as [206, Theorem 3]

$$g^{(1)}(\lambda) \geq \log\left(2\sqrt{\mathrm{Var}\left[Z\right]}\right) = \frac{1}{2}\log\left(k+2\lambda\right) + \frac{3}{2}\log 2. \tag{5.100}$$

Finally, let us now provide a lower bound for $g^{(2)}(\lambda)$ in (5.99). This can be obtained by applying Jensen's inequality:

$$g^{(2)}(\lambda) \geq -\frac{1}{2}\log\mathbb{E}\left[Z\right] = -\frac{1}{2}\log\left(k+\lambda\right). \tag{5.101}$$

Combining (5.97), (5.100), and (5.101), and returning to the original notation, we obtain

$$g(x) \geq \frac{1}{2}\log\left(\frac{2x^2+n\sigma_N^2}{x^2+n\sigma_N^2}\right) - \log\sigma_N + 3\log 2. \tag{5.102}$$

This lower bound on $g(x)$ is bounded inside an arbitrary finite interval $x \in [0, \hat{x}]$. Thus, the function $g(x)$ in the integral (5.92) is uniformly bounded via (5.94) and (5.102) in $[0, \hat{x}]$. Since in the asymptotic limit $\sigma_S \to \infty$ one has $\eta \to 0$ and $p_X(x) \to 0$ from (5.92), it follows that (5.76) is fulfilled as well, which concludes the proof.

5.9 Summary

The channel law for amplitude-modulated solitons transmitted through an NL optical fibre with ideal distributed amplification and a receiver based on the forward NFT is a noncentral chi-distribution with $2n$ degrees of freedom, where $n = 2$ and $n = 3$ correspond to the single- and dual-polarisation cases, respectively. For $n = 2$ case, an exact closed-form expression for a lower bound on the capacity of the NL optical fibre channel with no in-line dispersion compensation was derived. An asymptotic semi-analytic approximation for a capacity lower bound for arbitrary n and a Rayleigh input distribution was developed. It was shown that the lower bounds grows logarithmically with effective SNR, irrespective to the value of n. Numerical results for other continuous input distributions are also provided. A half-Gaussian input distribution is shown to give larger rates than a Rayleigh input distribution for $n = 1, 2, 3$. At an SNR of 25 dB, the best lower bounds we developed are approximately 3.68 bit per channel use. The practically relevant case of ASK constellations is also numerically analysed. For the same SNR of 25 dB, a 16-ASK constellation yields a rate of approximately 3.45 bit per channel use.

Chapter 6

General Conclusions

We review now the main contributions of this thesis.

Higher information rates are essential to cope with the ever-growing data demand. Suppressing the impact of optical Kerr nonlinearity in a fibre becomes one of the main challenges for enhancing the achievable information rates of optical fibre communication systems. The nonlinear interference process gives rise to the noise-like distortions in multi-channel WDM optical fibre communication systems, which severely constrains the achievable capacity. All existing nonlinearity compensation methods operated over the entire signal bandwidth can completely suppress deterministic signal-signal interactions, however, the nonlinear distortions due to interactions between signal and noise are left uncompensated because of their stochastic nature. This sets certain limitations on the performance of nonlinearity-compensated optical fibre communication systems. This also leads to an increased optimum transmitted power in comparison with traditional receiver schemes. In this thesis, we proposed an input-dependent analytical channel model for nonlinearity-compensated optical fibre communication systems considering the impact of both first- and second-order contributions of signal-noise nonlinear mixing. The proposed analytical model is successfully verified for different signal modulation formats via numerical simulations for ultra-wide transmitted bandwidths up to 1 THz. It should be noted that this analytical approach also allows us to accurately predict the performance of ultra-wide bandwidth multi-channel transmission systems with full-field nonlinearity compensation at optimum launched powers and beyond without using time-consuming and sometimes even intractable numerical simulations.

It should also be emphasised that the numerical simulations of ultra-wide bandwidth communication systems beyond 1 THz becomes infeasible, which, in fact, is a cogent reason behind developing powerful and effective analytical

methods. In this thesis, the limits on the achievable information rates of ultra-wideband optical communication systems using 40 nm erbium-doped fibre amplifier and 100 nm non-ideal distributed Raman amplification are estimated based on the verified first-order perturbative model at optimum power regime. In order to increase accuracy, we avoided the assumption of the "whiteness" of the nonlinear distortion spectra by means of evaluating the nonlinear interference for in each individual channel. Higher modulation formats together with probabilistically-shaped signal constellation were considered to approach the estimated theoretical limits.

On the other hand, it is now widely accepted that the nonlinear optical fibre channel capacity remains hugely unexplored. In the second part of the thesis, we considered the use of optical solitons to encode, transmit and decode information embedded into the soliton amplitudes, that is, imaginary part of the nonlinear discrete spectrum. Generally speaking, optical soliton pulses represent the "natural" solutions of lossless and noiseless NLSE, moreover, as opposed to any other waveforms, they are able to preserve its spectrum upon nonlinear transmission. However, the traditional soliton based transmission systems (*i.e.,* the conventional OOK modulation format), which employ the energy detection, has a very low spectral efficiency of 0.2 bit/s/Hz, and severely limited by the Gordon-Haus jitter effect, as well as the inter-soliton interaction between the WDM soliton channels. The use of NFT-based detection approaches is expected to be much more powerful and promising than traditional OOK. For such schemes, a non-Gaussian channel model for the conditional PDF of well-separated (in time) noise-perturbed soliton amplitudes, addressing the continuous modulation of the discrete eigenvalue associated with the NLSE, was derived and used to study lower bounds on the channel capacity. Results for propagation of signals over a NL optical fibre using one and two polarisations were also presented. The results demonstrated both analytically and numerically that there exist lower bounds on the channel capacity that display an unbounded growth with the effective SNR, similar to the linear Gaussian channel. However, it is aslo worth mentioning that all results are given in bit per channel use only, and thus, they should be considered as a first step towards analysing the more practically relevant problem of channel capacity in bit per second per unit bandwidth, *i.e.,* the spectral efficiency. The proper definition of spectral efficiency of optical soliton based communication systems remains a "subtle" problem since optical solitons require an infinite bandwidth. In order to increase the achievable spectral efficiency given in bits per second per unit bandwidth, in practice, the exponentially decaying tails of

solitons are supposed to be truncated before propagation through optical fibre. This, in turn, changes the NFT spectrum and breaks down the integrability of NLSE. These considerably more challenging problems are left for further investigation. Importantly, even though optical solitons are able to handle inherent fibre Kerr nonlinearity, the obtained spectral efficiencies are, so far, not even competitive with the traditional Nyquist-spaced communication systems, which is mainly due to relatively large time-bandwidth products.

Apart from the vector soliton channel model we also studied lower bounds on the capacity of an abstract generalised noncentral chi-channel with an arbitrary number of degrees of freedom. Similar channel models appear in the study of relatively general systems of noise-driven coupled nonlinear oscillators [207]. Therefore, we believe that the results for large number of degrees of freedom might also some day find applications in nonlinear communication channels.

The obtained results for the general noncentral chi-channel are true capacity lower bounds for that channel model. For the case of the considered application (*e.g.,* amplitude-modulated soliton systems), however, the presented analysis was based on a perturbative-based model which holds at high SNR regime. This model also does not consider potential interaction between solitons, and thus, the results are limited to solitons well-separated in a time domain. Another way of interpreting these results is that the obtained expressions are approximate lower bounds on the capacity of the true channel. Bounds that consider memory effects are left for further investigation. Furthermore, another interesting open research problem is the derivation of capacity upper bounds for amplitude-modulated soliton systems. This is also left for further investigation.

To summarise, in spite of the fact that the current performance of both soliton and NFT based transmissions is still incomparable with traditional communication systems, we do believe that these mathematically elegant approaches, which are entirely suited for nonlinear systems, can pave the way to the development of drastically non-conventional techniques for coding, modulation, transmission and signal processing for inherently nonlinear optical fibre communication systems.

Appendix A

Fokker-Plank equation

Here we give a concise description of the Fokker-Plank equation approach to derive the non-Gaussian channel law of the NL optical fibre channel in the presence of additive white Gaussian noise arising from optical amplification, which were used in Chapter 5.6 (see Sec. 5.6).

Owing to the low value of D in Eq. (5.7), we can employ adiabatic soliton perturbation theory [179, 208, 209]. Therefore, the scalar soliton parameters Eqs (5.25)-(5.28) must acquire noisy perturbative terms, which are well-described in [83]. Omitting some bulky analytical derivations (provided in, *e.g.,* [179]) and assuming a large propagation distance, we can obtain the following autonomous Fokker-Planck equation for a joint PDF of soliton frequency Ω and soliton amplitude A in Eq. (5.23). Hence, we have

$$\frac{\partial p_{\Omega,A}(\omega,a\,;\,z)}{\partial z} = \frac{D}{2}\,a\,\frac{\partial^2 p_{\Omega,A}(\omega,a\,;\,z)}{\partial a^2} + \frac{D}{6}\,a\,\frac{\partial^2 p_{\Omega,A}(\omega,a\,;\,z)}{\partial \omega^2}\,. \tag{A.1}$$

Eq. (A.1) can be solved by applying the Fourier transform with respect to Ω and the Laplace transform with respect to A, therefore, we have

$$\tilde{\tilde{p}}_{Z,\Xi}\,(\zeta,\xi\,;\,z) = \int\limits_{-\infty}^{\infty} d\omega\, e^{\imath\xi\omega} \int\limits_{0}^{\infty} da\, p_{\Omega,A}(\omega,a\,;\,z)\, e^{-\zeta a}\,. \tag{A.2}$$

After this transformation, Eq. (A.1) yields

$$\frac{\partial \tilde{\tilde{p}}_{Z,\Xi}\,(\zeta,\xi\,;\,z)}{\partial z} = D\left(\frac{m^2}{6} - \frac{\zeta^2}{2}\right)\frac{\partial \tilde{\tilde{p}}_{Z,\Xi}\,(\zeta,\xi\,;\,z)}{\partial \zeta} - D\zeta\,\tilde{\tilde{p}}_{Z,\Xi}\,(\zeta,\xi\,;\,z)\,. \tag{A.3}$$

Eq. (A.3) can be solved by means of the method of characteristics with the initial condition $\tilde{\tilde{p}}_{Z,\Xi}\,(\zeta,\xi\,;\,z)\big|_{z=0} = \exp\left(-\zeta a_0\right)$, where a_0 is the initial (input) value of a scalar soliton amplitude. Thus, the solution of Eq. (A.1) can be

expressed as follows

$$p_{\Omega,A}(\omega,a\,;\,z)=\frac{1}{2\pi}\int\limits_{-\infty}^{\infty}d\xi\,\tilde{p}_{A,\Xi}\left(a,\xi\,;\,z'\right)e^{-\imath\xi\omega}\,, \tag{A.4}$$

where

$$\begin{aligned}\tilde{p}_{A,\Xi}\left(a,\xi\,;\,z'\right)=&\left(\frac{a}{a_0}\right)^{1/2}\frac{\xi}{\sqrt{3}\sinh\left(\frac{\xi}{\sqrt{3}}z'\right)}\\&\times\exp\left[-\frac{\xi}{\sqrt{3}}\left(a_0+a\right)\coth\left(\frac{\xi}{\sqrt{3}}z'\right)\right]I_1\left(\frac{2\xi\sqrt{a\,a_0}}{\sqrt{3}\sinh\left(\frac{\xi}{\sqrt{3}}z'\right)}\right),\end{aligned} \tag{A.5}$$

where $z'\triangleq\frac{D}{2}\cdot z$, and $\coth(\cdot)$ is the hyperbolic cotangent. By taking the integral Eq. (A.4), we obtain the marginal conditional PDF for soliton amplitudes[1]

$$p_{A|A_0}\left(a|a_0\,;\,z'\right)=\frac{1}{z'}\sqrt{\frac{a}{a_0}}\exp\left(-\frac{a_0+a}{z'}\right)I_1\left(\frac{2\sqrt{a_0\,a}}{z'}\right). \tag{A.6}$$

Importantly, Eq. (A.6) represents the noncentral chi-squared distribution with four degrees of freedom and non-centrality parameter $4a_0/(z')^2$.

[1]Expression (A.6) is, in fact, the same PDF obtained assuming an energy-detection receiver (*i.e.,* a receiver based on Eq. (5.43)), as shown in [83, eq. (5.501)].

Appendix B

Fourier collocation method

Here we briefly describe the main steps of the so-called Fourier collocation method that has been used to numerically evaluate the forward NFT described in Sec. 5.3. The linear Zakharov-Shabat eigenvalue problem Eq. (5.9) for the NLSE Eq. (5.5) can be represented as follows[1]

$$\begin{pmatrix} -\frac{\partial}{\partial t} & \imath\, q(z,t) \\ -\imath\, q^*(z,t) & \frac{\partial}{\partial t} \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \imath\, \lambda \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}. \tag{B.1}$$

The key idea of the method relies on taking the discrete Fourier transform of Eq. (B.1), *i.e.,* taking the discrete Fourier transform of the sampled Jost functions $\Psi_{j,1}[k] \triangleq \Psi_1(z, mT; \lambda_j)$, and $\Psi_{j,2}[k] \triangleq \Psi_1(z, mT; \lambda_j)$ and sampled input potential (waveform) $q_0[k] \triangleq q(0, kT)$, where T denotes the sampling period. Thus, the discrete Fourier transforms with number of modes $M = 2N+1$, where N represents the number of eigenvalues for a given waveform, are given by

$$\Psi_{j,1}[k] = \sum_{m=-N}^{N} \alpha_j[m] \exp\left(\imath \frac{2\pi\, m}{M} k\right), \tag{B.2}$$

$$\Psi_{j,2}[k] = \sum_{m=-N}^{N} \beta_j[m] \exp\left(\imath \frac{2\pi\, m}{M} k\right), \tag{B.3}$$

$$q_0[k] = \sum_{m=-N}^{N} \gamma[m] \exp\left(\imath \frac{2\pi m}{M} k\right). \tag{B.4}$$

In order to compute the appropriate discrete Fourier amplitudes (coefficients) $\{\alpha_j[m], \beta_j[m]\}$, we substitute Eqs (B.2) (B.3) (B.4) into Eq. (B.1), we then obtain Zakharov-Shabat system Eq. (B.1), which are written in terms of the

[1]Here we suppose the absence of continuous spectrum, *i.e.,* $\mathscr{Q}_{\mathrm{c}}(\zeta) = 0$.

discrete Fourier transform, we have

$$\hat{\mathcal{H}} \begin{bmatrix} \boldsymbol{A}_j \\ \boldsymbol{B}_j \end{bmatrix} = \lambda_j \begin{bmatrix} \boldsymbol{A}_j \\ \boldsymbol{B}_j \end{bmatrix}, \tag{B.5}$$

where the non-Hermitian matrix $\hat{\mathcal{H}}$ is defined as

$$\hat{\mathcal{H}} \triangleq \begin{pmatrix} -\hat{\mathcal{C}} & \hat{\mathcal{T}} \\ -\hat{\mathcal{T}}^{\dagger} & \hat{\mathcal{C}} \end{pmatrix}, \tag{B.6}$$

where "†" stands for the Hermitian transpose, and $\hat{\mathcal{C}}$ is the following time-differential diagonal matrix

$$\hat{\mathcal{C}} \triangleq \frac{2\pi m}{M} \operatorname{diag}\left[-N, -N+1, \ldots, N-1, N\right], \tag{B.7}$$

and $\hat{\mathcal{T}} \in \mathbb{C}^{M \times M}$ is the following convolution Toeplitz matrix

$$\hat{\mathcal{T}} \triangleq \begin{pmatrix} \gamma[0] & \gamma[-1] & \ldots & \gamma[-N] & 0 & \ldots & 0 \\ \gamma[1] & \gamma[0] & \gamma[-1] & \ddots & \ddots & \ddots & \vdots \\ \vdots & \gamma[1] & \gamma[0] & \ddots & \ddots & \ddots & 0 \\ \gamma[N] & \ldots & \ldots & \ldots & \ldots & \ldots & \gamma[-N] \\ 0 & \gamma[N] & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \gamma[-1] \\ 0 & \ldots & 0 & \gamma[N] & \ldots & \gamma[1] & \gamma[0] \end{pmatrix}, \tag{B.8}$$

and the discrete Fourier vector amplitudes are

$$\boldsymbol{A}_j \triangleq \left[\alpha_j[-N], \alpha_j[-N+1], \ldots, \alpha_j[N-1], \alpha_j[N]\right]^{\mathrm{T}}, \tag{B.9}$$

$$\boldsymbol{B}_j \triangleq \left[\beta_j[-N], \beta_j[-N+1], \ldots, \beta_j[N-1], \beta_j[N]\right]^{\mathrm{T}}. \tag{B.10}$$

For relatively small N the eigenvalues can be computed by using the well-known QR-algorithm (see, *e.g.,* [210]), whereas for a large N it can be efficiently evaluated via the Arnoldi algorithm (see, *e.g.,* [211, 212])[2].

[2]Both these algorithms have already implemented in MatLab.

Appendix C

Proof of Lemmas 5.6.1 and 5.6.2

Proof of Lemma 5.6.1. As a first step, we shall calculate the output distribution function, which is given by

$$p_Y(y) = \int_0^\infty p_{Y|X}(y|x)\, p_X(x)\, dx. \tag{C.1}$$

Substituting Eq. (5.38) and Eq. (5.44) into Eq. (C.1), we obtain

$$p_Y(y) \triangleq \frac{4y^2}{\sigma_S^2\sigma_N^2} \exp\left(-\frac{y^2}{\sigma_N^2}\right) \int_0^\infty \exp\left(-\frac{\sigma_S^2+\sigma_N^2}{\sigma_S^2\sigma_N^2}x^2\right) I_1\left(\frac{2y}{\sigma_N^2}x\right) dx \tag{C.2}$$

In order to take the integral (C.2) we should expand the modified Bessel function $I_1(\cdot)$ into a power series, and then change the order of summation and integration. Thus, the output distribution reads

$$p_Y(y) = 2\sqrt{\frac{\pi}{\sigma_S^2\sigma_N^2(\sigma_S^2+\sigma_N^2)}}\, y^2 \exp\left(-\alpha y^2\right) I_{1/2}\left(\beta y^2\right) \tag{C.3}$$

$$= \frac{4}{\sigma_S^2}\, y \exp\left(-\alpha y^2\right) \sinh\left(\beta y^2\right), \tag{C.4}$$

where we have introduced the following notations

$$\begin{aligned} \alpha &\triangleq \frac{\sigma_S^2+2\sigma_N^2}{2\sigma_N^2(\sigma_S^2+\sigma_N^2)} \\ \beta &\triangleq \frac{\sigma_S^2}{2\sigma_N^2(\sigma_S^2+\sigma_N^2)} \end{aligned} \tag{C.5}$$

and $I_{1/2}(x) \triangleq \sqrt{2/(\pi x)}\sinh(x)$ denotes the half-order modified Bessel function of the first kind, and $\sinh(x)$ denotes a hyperbolic sine. We now express the integral of the output differential entropy for the output distribution function,

i.e.,

$$h_Y \triangleq -\int_0^\infty p_Y(y)\log p_Y(y)\,dy \tag{C.6}$$

as a sum of four integrals by splitting the logarithm of (C.3) as follows

$$\log p_{\rm Y}(y) = \log\frac{4}{\sigma_S^2} - \alpha y^2 + \log y + \log\left[\sinh\left(\beta y^2\right)\right], \tag{C.7}$$

and then compute the contribution of each term separately. Thus, the differential output entropy can be represented as follows: $h_Y = h_Y^{(1)} + h_Y^{(2)} + h_Y^{(3)} + h_Y^{(4)}$.

The first contribution is given by the following integral [117, Eq. (3.541.1)]:

$$h_Y^{(1)} = -\frac{4}{\sigma_S^2}\log\frac{4}{\sigma_S^2}\int_0^\infty y\exp\left(-\alpha y^2\right)\sinh\left(\beta y^2\right)\,dy \tag{C.8}$$

$$= \frac{2}{\sigma_S^2}\log\frac{\sigma_S^2}{4}\int_0^\infty \exp(-\alpha t)\sinh(\beta t)\,dt \tag{C.9}$$

$$= \frac{\sigma_N^2}{\sigma_S^2}\left(1+\frac{\sigma_N^2}{\sigma_S^2}\right)\mathrm{B}\left(\frac{\sigma_N^2}{\sigma_S^2},2\right)\log\frac{\sigma_S^2}{4} \tag{C.10}$$

$$= \log\sigma_S^2 - 2\log 2, \tag{C.11}$$

where $\mathrm{B}(x,y) \triangleq \int_0^\infty (1+t)^{-(x+y)}t^{x-1}\,dt$ stands for the Beta function (Euler's integral of the first kind).

The second integral can be taken using integration by parts:

$$h_Y^{(2)} = \frac{4\alpha}{\sigma_S^2}\int_0^\infty y^3\exp\left(-\alpha y^2\right)\sinh\left(\beta y^2\right)\,dy \tag{C.12}$$

$$= \frac{2\alpha}{\sigma_S^2}\int_0^\infty t\exp(-\alpha t)\sinh(\beta t)\,dt \tag{C.13}$$

$$= \frac{\alpha}{\sigma_S^2}\left\{\int_0^\infty t\exp\left[-(\alpha-\beta)t\right]\,dt - \int_0^\infty t\exp\left[-(\alpha+\beta)t\right]\,dt\right\} \tag{C.14}$$

$$= \frac{\alpha}{\sigma_S^2}\left[\frac{1}{(\alpha-\beta)^2} - \frac{1}{(\alpha+\beta)^2}\right] \tag{C.15}$$

$$= \frac{(\sigma_S^2+2\sigma_N^2)^2}{2\sigma_N^2(\sigma_S^2+\sigma_N^2)}. \tag{C.16}$$

The third one is obtained as

$$h_Y^{(3)} = -\frac{4}{\sigma_S^2}\int_0^\infty y\exp\left(-\alpha y^2\right)\sinh\left(\beta y^2\right)\log y\,dy \tag{C.17}$$

$$= -\frac{1}{\sigma_S^2}\int_0^\infty \exp(-\alpha t)\sinh(\beta t)\log t\,dt \tag{C.18}$$

$$= \frac{1}{2\sigma_S^2}\Bigg\{\int_0^\infty \exp\left[-(\alpha+\beta)t\right]\log t\,dt$$

$$-\int_0^\infty \exp\left[-(\alpha-\beta)t\right]\log t\,dt\Bigg\} \tag{C.19}$$

$$= \frac{\gamma}{2} - \log\sqrt{\sigma_S^2+\sigma_N^2} - \frac{\sigma_N^2}{\sigma_S^2}\log\sqrt{1+\frac{\sigma_S^2}{\sigma_N^2}}, \tag{C.20}$$

where (C.20) follows from [117, Eq. (4.331.1)].

Finally, the forth term can be expressed in closed form as follows

$$h_{\mathrm{Y}}^{(4)} = -\frac{4}{\sigma_S^2}\int_0^\infty y\exp\left(-\alpha y^2\right)\sinh\left(\beta y^2\right)\log\left[\sinh\left(\beta y^2\right)\right]dy \tag{C.21}$$

$$= -\frac{2}{\sigma_S^2}\int_0^\infty \exp(-\alpha t)\sinh(\beta t)\log\left[\sinh(\beta t)\right]dt \tag{C.22}$$

$$= \frac{1}{\sigma_S^2\beta}\Bigg\{\int_0^\infty \exp\left[-\left(\frac{\alpha+\beta}{\beta}\right)t\right]\log\left[\sinh(t)\right]dt$$

$$-\int_0^\infty \exp\left[-\left(\frac{\alpha-\beta}{\beta}\right)t\right]\log\left[\sinh(t)\right]dt\Bigg\} \tag{C.23}$$

$$= -\frac{1}{\sigma_S^2}\frac{2\beta}{(\alpha^2-\beta^2)}\Bigg\{\frac{2\alpha\beta}{\alpha^2-\beta^2} - \log 2$$

$$+\frac{1}{2\beta}\left[(\alpha-\beta)\cdot H_{\frac{\alpha+\beta}{2\beta}} - \left(\alpha+\beta\right)\cdot H_{\frac{\alpha-\beta}{2\beta}}\right]\Bigg\} \tag{C.24}$$

$$= \frac{\sigma_S^4-2\sigma_N^4}{2\sigma_N^2(\sigma_S^2+\sigma_N^2)} + \psi\left(\frac{\sigma_N^2}{\sigma_S^2}\right) + \gamma + \log 2, \tag{C.25}$$

where $H_n \triangleq \sum_{l=1}^{n}\frac{1}{l}$ is the n-th harmonic number. To pass from (C.23) to (C.24) we used [213, Eq. (256.14)].

The proof of the theorem is completed by combining all these terms: (C.11), (C.16), (C.20), and (C.25). ■

Proof of Lemma 5.6.2. Similar to the previous case, we represent the logarithm of a product as a sum of logarithms of the factors, each of which gives their own contribution to the integral, *i.e.,*

$$\log\Big[p_{Y|X}(y|x)\Big] = \log\left(\frac{2}{\sigma_N^2}\frac{1}{x}e^{-x^2/\sigma_N^2}\right) - \frac{y^2}{\sigma_N^2} + 2\log y + \log\left[I_1\left(\frac{2xy}{\sigma_N^2}\right)\right]. \quad \text{(C.26)}$$

Therefore, the differential conditional entropy defined as the following improper double integral

$$h_{Y|X} \triangleq -\int_0^\infty\int_0^\infty p_X(x)p_{Y|X}(y|x)\log p_{Y|X}(y|x)\,dxdy \quad \text{(C.27)}$$

can be split into four different terms:

$$h_{Y|X} = h_{Y|X}^{(1)} + h_{Y|X}^{(2)} + h_{Y|X}^{(3)} + h_{Y|X}^{(4)}. \quad \text{(C.28)}$$

The first term gives

$$h_{Y|X}^{(1)} = \frac{-4}{\sigma_S^2\sigma_N^2}\int_0^\infty \exp\left[-\left(\frac{\sigma_S^2+\sigma_N^2}{\sigma_S^2\sigma_N^2}\right)x^2\right]\log\left(\frac{2e^{-x^2/\sigma_N^2}}{x\sigma_N^2}\right)\mathcal{I}^{(1)}(x)\,dx \quad \text{(C.29)}$$

where the integral $\mathcal{I}^{(1)}(x)$ over y is a convergent improper integral and can be taken by using exactly the same approach as the integral (C.2), we then have

$$\mathcal{I}^{(1)}(x) = \int_0^\infty y^2\exp\left(-\frac{y^2}{\sigma_N^2}\right)I_1\left(\frac{2xy}{\sigma_N^2}\right)dy \quad \text{(C.30)}$$

$$= \int_0^\infty y^2\exp\Big(-p^2y^2\Big)I_1(by)\,dy \quad \text{(C.31)}$$

$$= \sum_{k=0}^\infty \frac{1}{k!(k+1)!}\left(\frac{b}{2}\right)^{2k+1}\int_0^\infty y^{2k+3}\exp\Big(-p^2y^2\Big)\,dy \quad \text{(C.32)}$$

$$= \frac{b}{(2p^2)^2}\sum_{k=0}^\infty \frac{1}{k!}\left(\frac{b}{2p}\right)^{2k} \quad \text{(C.33)}$$

$$= \frac{b^2}{(2p^2)^2}\exp\left(\frac{b^2}{4p^2}\right) = \frac{\sigma_N^2}{2}x\exp\left(\frac{x^2}{\sigma_N^2}\right), \quad \text{(C.34)}$$

where to pass from (C.32) to (C.33) we use [117, Eq. (3.462.9)]. Thus, we have

$$h_{Y|X}^{(1)} = -\frac{2}{\sigma_S^2}\int_0^\infty x e^{-x^2/\sigma_S^2} \log\left(\frac{2e^{-x^2/\sigma_N^2}}{x\sigma_N^2}\right) dx \tag{C.35}$$

$$= h_{Y|X}^{(1,1)} + h_{Y|X}^{(1,2)} + h_{Y|X}^{(1,3)}, \tag{C.36}$$

where

$$h_{Y|X}^{(1,1)} = \frac{2}{\sigma_S^2}\log\left(\frac{\sigma_N^2}{2}\right)\int_0^\infty x\exp\left(-\frac{x^2}{\sigma_S^2}\right) dx \tag{C.37}$$

$$= \frac{2}{\sigma_S^2}\log\left(\frac{\sigma_N^2}{2}\right)\int_0^\infty x\exp\left(-q^2x^2\right) dx \tag{C.38}$$

$$= \frac{2}{\sigma_S^2}\log\left(\frac{\sigma_N^2}{2}\right)\frac{1}{2q^2} = \log\frac{\sigma_N^2}{2}. \tag{C.39}$$

$$h_{Y|X}^{(1,2)} = \frac{2}{\sigma_S^2}\int_0^\infty x\exp\left(-\frac{x^2}{\sigma_S^2}\right)\log x\, dx \tag{C.40}$$

$$= \frac{2}{\sigma_S^2}\int_0^\infty x\exp\left(-q^2x^2\right)\log x\, dx \tag{C.41}$$

$$= \frac{1}{2\sigma_S^2}\int_0^\infty \exp\left(-q^2t\right)\log t\, dt \tag{C.42}$$

$$= -\frac{1}{2\sigma_S^2}\frac{1}{q^2}\left(\gamma + 2\log q\right) \tag{C.43}$$

$$= \frac{1}{2}\log\sigma_S^2 - \frac{\gamma}{2}, \tag{C.44}$$

where (C.43) follows from [117, Eq. (4.331.1)].

$$h_{Y|X}^{(1,3)} = \frac{2}{\sigma_S^2\sigma_N^2}\int_0^\infty x^3\exp\left(-\frac{x^2}{\sigma_S^2}\right) dx \tag{C.45}$$

$$= \frac{2}{\sigma_S^2\sigma_N^2}\int_0^\infty x^3\exp\left(-q^2x^2\right) dx \tag{C.46}$$

$$= \frac{2}{\sigma_S^2\sigma_N^2}\frac{1}{2q^4} = \frac{\sigma_S^2}{\sigma_N^2}. \tag{C.47}$$

To sum up all these contributions the first term of conditional entropy reads

$$h_{Y|X}^{(1)} = \log\frac{\sigma_N^2}{2} + \frac{1}{2}\log\sigma_S^2 + \frac{\sigma_S^2}{\sigma_N^2} - \frac{\gamma}{2}. \tag{C.48}$$

The second term is

$$h_{Y|X}^{(2)} = -\frac{4}{\sigma_S^2\sigma_N^2}\int_0^\infty \exp\left[-\left(\frac{1}{\sigma_S^2}+\frac{1}{\sigma_N^2}\right)x^2\right]\mathcal{I}^{(2)}(x)\,dx, \tag{C.49}$$

where the integral $\mathcal{I}^{(2)}(x)$ is given by

$$\mathcal{I}^{(2)}(x) = -\frac{1}{\sigma_N^2}\int_0^\infty y^4 \exp\left(-\frac{y^2}{\sigma_N^2}\right) I_1\left(\frac{2xy}{\sigma_N^2}\right) dy \tag{C.50}$$

$$= -\frac{1}{\sigma_N^2}\int_0^\infty y^4 \exp\left(-p^2y^2\right) I_1(by)\,dy \tag{C.51}$$

$$= -\frac{1}{\sigma_N^2}\sum_{k=0}^\infty \frac{1}{k!(k+1)!}\left(\frac{b}{2}\right)^{2k+1}\int_0^\infty y^{2k+5}\exp\left(-p^2y^2\right) dy \tag{C.52}$$

$$= -\frac{1}{\sigma_N^2}\frac{b}{4(p^2)^3}\sum_{k=0}^\infty \frac{(k+2)}{k!}\left(\frac{b}{2p}\right)^{2k} \tag{C.53}$$

$$= -\frac{1}{\sigma_N^2}\frac{b}{4(p^2)^3}\left(\frac{b^2}{4p^2}+2\right)\exp\left(\frac{b^2}{4p^2}\right) \tag{C.54}$$

$$= -\frac{1}{2}x\left(x^2+2\sigma_N^2\right)\exp\left(\frac{x^2}{\sigma_N^2}\right). \tag{C.55}$$

We have

$$h_{Y|X}^{(2)} = \frac{2}{\sigma_S^2\sigma_N^2}\int_0^\infty x\left(x^2+2\sigma_N^2\right)\exp\left(-\frac{x^2}{\sigma_S^2}\right) dx = h_{Y|X}^{(2,1)} + h_{Y|X}^{(2,2)}, \tag{C.56}$$

where

$$h_{\mathrm{Y|X}}^{(2,1)} = \frac{2}{\sigma_S^2\sigma_N^2}\int_0^\infty x^3\exp\left(-\frac{x^2}{\sigma_S^2}\right) dx \tag{C.57}$$

$$= \frac{2}{\sigma_S^2\sigma_N^2}\int_0^\infty x^3\exp\left(-q^2x^2\right) dx \tag{C.58}$$

$$= \frac{2}{\sigma_S^2\sigma_N^2}\frac{1}{2q^4} = \frac{\sigma_S^2}{\sigma_N^2}. \tag{C.59}$$

$$h_{Y|X}^{(2,2)} = \frac{4}{\sigma_S^2}\int_0^\infty x\exp\left(-\frac{x^2}{\sigma_S^2}\right)dx \tag{C.60}$$

$$= \frac{4}{\sigma_S^2}\int_0^\infty x\exp\left(-q^2x^2\right)dx \tag{C.61}$$

$$= \frac{4}{\sigma_S^2}\frac{1}{2q^2} = 2. \tag{C.62}$$

The second term of conditional entropy $h_{Y|X}^{(2)}$ reads

$$h_{Y|X}^{(2)} = \frac{\sigma_S^2}{\sigma_N^2} + 2. \tag{C.63}$$

The third term is

$$h_{Y|X}^{(3)} = -\frac{8}{\sigma_S^2\sigma_N^2}\int_0^\infty \exp\left[-\left(\frac{1}{\sigma_S^2}+\frac{1}{\sigma_N^2}\right)x^2\right]\mathcal{I}^{(3)}(x)\,dx \tag{C.64}$$

where

$$\mathcal{I}^{(3)}(x) = \int_0^\infty y^2\exp\left(-\frac{y^2}{\sigma_N^2}\right)I_1\left(\frac{2xy}{\sigma_N^2}\right)\log y\,dy \tag{C.65}$$

$$= \int_0^\infty y^2\exp\left(-p^2y^2\right)I_1(by)\log y\,dy \tag{C.66}$$

$$= \sum_{k=0}^\infty \frac{1}{k!(k+1)!}\left(\frac{b}{2}\right)^{2k+1}\int_0^\infty y^{2k+3}\exp\left(-p^2y^2\right)\log y\,dy \tag{C.67}$$

$$= \frac{b}{2^3}\sum_{k=0}^\infty \frac{1}{k!(k+1)!}\left(\frac{b}{2}\right)^{2k}\int_0^\infty t^{k+1}\exp\left(-p^2t\right)\log t\,dt \tag{C.68}$$

$$= \frac{b}{2^3}\sum_{k=0}^\infty \frac{1}{(p^2)^{k+2}k!}\left[\psi(k+2)-\log p^2\right]\left(\frac{b}{2}\right)^{2k} \tag{C.69}$$

$$= \frac{b}{2^3(p^2)^2}\sum_{k=0}^\infty \frac{1}{k!}\left[\psi(k+1)+\frac{1}{k+1}-\log p^2\right]\left(\frac{b}{2p}\right)^{2k} \tag{C.70}$$

$$= \frac{b}{2^3(p^2)^2}\left\{\sum_{k=0}^\infty \frac{\psi(k+1)}{k!}\left(\frac{b}{2p}\right)^{2k}\right.$$
$$\left.+\sum_{k=0}^\infty \frac{1}{(k+1)k!}\left(\frac{b}{2p}\right)^{2k} - \log\left(p^2\right)\sum_{k=0}^\infty \frac{1}{k!}\left(\frac{b}{2p}\right)^{2k}\right\} \tag{C.71}$$

$$= \frac{1}{2bp^2}\left[\exp\left(\frac{b^2}{4p^2}\right)-1\right]$$

$$+\frac{b}{2^3(p^2)^2}\left[\Gamma\left(0,\frac{b^2}{4p^2}\right)+\log\left(\frac{b^2}{4(p^2)^2}\right)\right]\exp\left(\frac{b^2}{4p^2}\right) \tag{C.72}$$

$$=\frac{\sigma_{\mathrm{N}}^4}{4}\frac{1}{x}\left[\exp\left(\frac{x^2}{\sigma_N^2}\right)-1\right]$$

$$+\frac{\sigma_N^2}{4}x\left[2\log(x)-\mathrm{Ei}\left(-\frac{x^2}{\sigma_N^2}\right)\right]\exp\left(\frac{x^2}{\sigma_N^2}\right), \tag{C.73}$$

where $\Gamma(\alpha,x)\triangleq\int_x^\infty\exp(-t)t^{\alpha-1}\,dt$ is the incomplete gamma function, $\mathrm{Ei}(x)\triangleq-\int_{-x}^\infty t^{-1}\exp(-t)\,dt$ is the exponential integral function, $\Gamma(0,x)\triangleq-\mathrm{Ei}(-x)$. To pass from (C.68) to (C.69), the expression [117, Eq. (4.352.1)] was used.

Therefore, the second contribution can be split into following three terms

$$h_{Y|X}^{(3)}=h_{Y|X}^{(3,1)}+h_{Y|X}^{(3,2)}+h_{Y|X}^{(3,3)}, \tag{C.74}$$

where

$$h_{Y|X}^{(3,1)}=\frac{2\sigma_N^2}{\sigma_S^2}\int_0^\infty\left\{\exp\left[-\left(\frac{1}{\sigma_S^2}+\frac{1}{\sigma_N^2}\right)x^2\right]-\exp\left(-\frac{x^2}{\sigma_S^2}\right)\right\}\frac{dx}{x} \tag{C.75}$$

$$=\frac{2\sigma_N^2}{\sigma_S^2}\int_0^\infty\left\{\exp\left(-q^2x^2\right)\left[\exp\left(-p^2x^2\right)-1\right]\right\}\frac{dx}{x} \tag{C.76}$$

$$=\frac{\sigma_N^2}{\sigma_S^2}\int_0^\infty\left\{\exp(-\mu t)\Big[\exp(-t)-1\Big]\right\}\frac{dt}{t} \tag{C.77}$$

$$=-\frac{\sigma_N^2}{\sigma_S^2}\log\left(1+\frac{1}{\mu}\right)=-\frac{\sigma_N^2}{\sigma_S^2}\log\left(1+\frac{\sigma_S^2}{\sigma_N^2}\right), \tag{C.78}$$

where [117, Eq. (3.411.19)] was used to pass from (C.77) to (C.78). The integral $h_{Y|X}^{(3,2)}$ is the same as that of (C.40), so we have

$$h_{Y|X}^{(3,2)}=-\frac{4}{\sigma_S^2}\int_0^\infty x\exp\left(-\frac{x^2}{\sigma_S^2}\right)\log x\,dx \tag{C.79}$$

$$=-\frac{4}{\sigma_S^2}\int_0^\infty x\exp\left(q^2x^2\right)\log x\,dx \tag{C.80}$$

$$=-\frac{1}{\sigma_S^2}\int_0^\infty\exp\left(-q^2t\right)\log t\,dt \tag{C.81}$$

$$=\frac{1}{\sigma_S^2}\frac{1}{q^2}\left(\gamma+2\log q\right)=\gamma-\log\sigma_S^2. \tag{C.82}$$

The integral $h_{Y|X}^{(3,3)}$ follows from [214, Eq. (4.2.3)]:

$$h_{Y|X}^{(3,3)} = \frac{2}{\sigma_S^2}\int_0^\infty x\exp\left(-\frac{x^2}{\sigma_S^2}\right)\mathrm{Ei}\left(-\frac{x^2}{\sigma_N^2}\right)dx \tag{C.83}$$

$$= \frac{2}{\sigma_S^2}\int_0^\infty x\exp\left(-q^2x^2\right)\mathrm{Ei}\left(-p^2x^2\right)dx \tag{C.84}$$

$$= \frac{1}{\sigma_S^2}\int_0^\infty \exp\left(-q^2t\right)\mathrm{Ei}\left(-p^2t\right)dt \tag{C.85}$$

$$-\frac{1}{q^2\sigma_S^2}\log\left(1+\frac{q^2}{p^2}\right) = -\log\left(1+\frac{\sigma_N^2}{\sigma_S^2}\right). \tag{C.86}$$

Therefore, the third term of conditional entropy $h_{Y|X}^{(3)}$ reads

$$h_{Y|X}^{(3)} = \gamma - \log\left(\sigma_S^2+\sigma_N^2\right) - \frac{\sigma_N^2}{\sigma_S^2}\log\left(1+\frac{\sigma_S^2}{\sigma_N^2}\right). \tag{C.87}$$

Finally, the last fourth term is

$$h_{Y|X}^{(4)} = -\frac{4}{\sigma_S^2\sigma_N^2}\int_0^\infty\int_0^\infty y^2\exp\left(-\frac{x^2}{\sigma_S^2}\right)\exp\left(-\frac{x^2+y^2}{\sigma_N^2}\right)I_1\left(\frac{2xy}{\sigma_N^2}\right)\log\left[I_1\left(\frac{2xy}{\sigma_N^2}\right)\right]dxdy \tag{C.88}$$

The following variable substitution is suggested

$$\begin{cases} \xi = 2xy/\sigma_N^2 \\ \eta = y^2/\sigma_N^2 \end{cases} \tag{C.89}$$

Its inverse Jacobian is

$$J^{-1} \triangleq \frac{\partial(\xi,\eta)}{\partial(x,y)} = \begin{vmatrix} \frac{2y}{\sigma_N^2} & \frac{2x}{\sigma_N^2} \\ 0 & \frac{2y}{\sigma_N^2} \end{vmatrix} = \frac{4y^2}{\sigma_N^4}$$

Changing the order of integration in the new variables one arrives at

$$h_{Y|X}^{(4)} = -\frac{\sigma_N^2}{\sigma_S^2}\int_0^\infty I_1(\xi)\log\left[I_1(\xi)\right]\mathcal{I}^{(4)}(\xi)\,d\xi, \tag{C.90}$$

where

$$\mathcal{I}^{(4)}(\xi) = \int_0^\infty \exp\left[-\eta - \frac{\xi^2}{4\eta}\left(1+\frac{\sigma_N^2}{\sigma_S^2}\right)\right] d\eta \tag{C.91}$$

$$= \int_0^\infty \exp\left[-\left(\frac{\zeta}{4\eta}+\eta\right)\right] d\eta \tag{C.92}$$

$$= \sqrt{\zeta}\, K_1\left(\sqrt{\zeta}\right) \tag{C.93}$$

$$= \sqrt{1+\frac{\sigma_N^2}{\sigma_S^2}}\,\xi\, K_1\left(\sqrt{1+\frac{\sigma_N^2}{\sigma_S^2}}\,\xi\right), \tag{C.94}$$

where (C.94) follows from [117, Eq. (3.324.1)]. Hence, the final expression of the fourth contribution of $h_{Y|X}$ reads:

$$h_{Y|X}^{(4)} = -\frac{\sigma_N^2}{\sigma_S^2}\sqrt{1+\frac{\sigma_N^2}{\sigma_S^2}}\int_0^\infty \xi\, K_1\left(\sqrt{1+\frac{\sigma_N^2}{\sigma_S^2}}\,\xi\right) I_1(\xi)\log\left[I_1(\xi)\right] d\xi \tag{C.95}$$

where $K_1(x)$ is the modified Bessel function of the second kind (Macdonald function) of order one. It is evident that the integral (C.95) cannot be solved in closed form.

Combining all these terms gives (5.46). ■

Appendix D

Proof of Lemmas 5.6.5 and 5.6.6

Proof of Lemma 5.6.5. The output distribution function is

$$\begin{aligned} P_Y(y) &= \int_0^\infty p_{Y|X}(y|x)\, p_X(x)\, dx \\ &= \frac{4y^3}{\sigma_S^2\sigma_N^2} \exp\left(-\frac{y^2}{\sigma_N^2}\right) \int_0^\infty \exp\left[-\left(\frac{\sigma_S^2+\sigma_N^2}{\sigma_S^2\sigma_N^2}\right)x^2\right] I_2\left(\frac{2y}{\sigma_N^2}x\right)\frac{dx}{x} \\ &= C(y)\int_0^\infty \frac{e^{-ax^2}}{x} I_2(bx)\, dx = C(y)\mathcal{I}(y), \end{aligned} \tag{D.1}$$

where we have introduced $C(y) \triangleq \frac{4y^3}{\sigma_S^2\sigma_N^2}\, e^{-\frac{y^2}{\sigma_N^2}}$, $a \triangleq \frac{\sigma_S^2+\sigma_N^2}{\sigma_S^2\sigma_N^2}$, and $b \triangleq \frac{2y}{\sigma_N^2}$.

The integral $\mathcal{I}(y)$ in Eq. (D.1) can be calculated as follows

$$\begin{aligned} \mathcal{I}(y) &\triangleq \int_0^\infty \frac{e^{-ax^2}}{x} I_2(bx)\, dx \\ &= \int_0^\infty \frac{e^{-ax^2}}{x} \sum_{k=0}^\infty \frac{1}{k!(k+2)!}\left(\frac{bx}{2}\right)^{2k+2} dx \\ &= \frac{b^2}{4}\sum_{k=0}^\infty \frac{1}{k!(k+2)!}\left(\frac{b}{2}\right)^{2k} \int_0^\infty x^{2k+1} e^{-ax^2}\, dx \end{aligned}$$

$$\begin{aligned} &= \frac{b^2}{8a}\sum_{k=0}^\infty \frac{1}{(k+2)!}\left(\frac{b^2}{4a}\right)^k \\ &= \frac{b^2}{8a}\left(-\frac{4a}{b^2} - \frac{16a^2}{b^4} + \frac{16a^2}{b^4}\, e^{\frac{b^2}{4a}}\right) \\ &= -\frac{2a}{b^2}\left(1 - e^{\frac{b^2}{4a}}\right) - \frac{1}{2}. \end{aligned} \tag{D.2}$$

To solve (D.2), the following function serious was used

$$\sum_{k=0}^{\infty} \frac{x^k}{(k+2)!} = \frac{e^x}{x^2} - \frac{1}{x} - \frac{1}{x^2}\,. \tag{D.3}$$

Therefore, the output distribution function $p_Y(y)$ is

$$\begin{aligned} p_Y(y) &= -\frac{2a\,C(y)}{b^2}\left(1 - e^{\frac{b^2}{4a}}\right) - \frac{C(y)}{2} \\ &= \frac{2(\sigma_S^2+\sigma_N^2)\,y}{\sigma_S^4}\left(e^{-\frac{y^2}{\sigma_S^2+\sigma_N^2}} - e^{\frac{y^2}{\sigma_N^2}}\right) - \frac{2y^3}{\sigma_S^2\sigma_N^2}e^{-\frac{y^2}{\sigma_N^2}} \end{aligned} \tag{D.4}$$

The differential Shannon entropy (2.1) of output given by distribution $p_Y(y)$ has no closed-form solution. However, it can be "bluntly" lower-bounded by using *Jensen's inequality* (see, for example, [32]), *i.e.,*

$$\mathbb{E}\,[\,g_Y(y)\,] \geq g_Y(\mu)\,, \tag{D.5}$$

where $\mu \triangleq \mathbb{E}\,[Y]$ is the mean value of the RV Y and $g_Y(y)$ is a convex function.

Applying inequality (D.5), we have

$$h_Y = -\int_0^{\infty} p_Y(y)\log p_Y(y)\,dy \geq g_Y(\mu)\,, \tag{D.6}$$

where the mean value μ is given by

$$\mu \triangleq \mathbb{E}\,[\,Y\,] = \int_0^{\infty} y\,p_Y(y)\,dy\,, \tag{D.7}$$

and the function $g_Y(y)$ is defined as

$$g_Y(y) \triangleq -\log p_Y(y)\,. \tag{D.8}$$

Note that, the proof of convexity of the distribution function $g_Y(y)$ is rather straightforward and omitted here. The mean value μ can be readily calculated

$$\mu \triangleq \int_0^\infty y\, p_Y(y)\, dy = \frac{2(\sigma_S^2+\sigma_N^2)}{\sigma_S^4}\left(\mathcal{I}^{(1)} - \mathcal{I}^{(2)}\right) - \frac{2}{\sigma_S^2\sigma_N^2}\mathcal{I}^{(3)}, \tag{D.9}$$

where $\mathcal{I}^{(1)}$, $\mathcal{I}^{(2)}$, and $\mathcal{I}^{(3)}$ are the following three simple integrals

$$\mathcal{I}^{(1)} \triangleq \int_0^\infty y^2 \exp\left(-\frac{y^2}{\sigma_S^2+\sigma_N^2}\right) dy = \frac{(\sigma_S^2+\sigma_N^2)^{3/2}}{2}\,\Gamma\left(\frac{3}{2}\right), \tag{D.10}$$

$$\mathcal{I}^{(2)} \triangleq \int_0^\infty y^2 \exp\left(-\frac{y^2}{\sigma_N^2}\right) dy = \frac{\sigma_N^3}{2}\,\Gamma\left(\frac{3}{2}\right), \tag{D.11}$$

$$\mathcal{I}^{(3)} \triangleq \int_0^\infty y^4 \exp\left(-\frac{y^2}{\sigma_N^2}\right) dy = \frac{\sigma_N^5}{2}\,\Gamma\left(\frac{5}{2}\right), \tag{D.12}$$

hence

$$\mu(\rho) = \frac{\sqrt{\pi}}{2\rho^2}\left[(1+\rho)^{\frac{5}{2}} - \frac{5\rho}{2} - 1\right]. \tag{D.13}$$

Finally, the entropy h_Y has the following lower bound

$$h_Y^{(\mathrm{LB})} = -\log\left\{\frac{2(\rho+1)\,\mu(\rho)}{\rho^2}\left[\exp\left(-\frac{\mu^2(\rho)}{\rho+1}\right) - \exp\left(-\mu^2(\rho)\right)\right] - \frac{2\mu^3(\rho)}{\rho}\exp\left(-\mu^2(\rho)\right)\right\}. \tag{D.14}$$

■

Proof of Lemma 5.6.6. The differential condition Shannon entropy is defined as

$$h_{Y|X} \triangleq -\int_0^\infty p_X(x)\,\mathbb{E}\left[\log p_{Y|X}(y|x)\middle| X=x\right] dx. \tag{D.15}$$

Be representing the logarithm of a product as a sum of the factors, we have

$$\begin{aligned}\log p_{Y|X}(y|x) &= \log\left[\frac{2}{\sigma_N^2}\frac{y^3}{x^2}\exp\left(-\frac{x^2+y^2}{\sigma_N^2}\right) I_2\left(\frac{2xy}{\sigma_N^2}\right)\right] \\ &= \log\left(\frac{2}{\sigma_N^2}\frac{e^{-\frac{x^2}{\sigma_N^2}}}{x^2}\right) + 3\log y - \frac{y^2}{\sigma_N^2} + \log\left[I_2\left(\frac{2xy}{\sigma_N^2}\right)\right], \end{aligned}\tag{D.16}$$

we then have four different contributions, *i.e.*,

$$h_{Y|X} = h_{Y|X}^{(1)} + h_{Y|X}^{(2)} + h_{Y|X}^{(3)} + h_{Y|X}^{(4)}. \tag{D.17}$$

The 1st-term $h_{Y|X}^{(1)}$ is captured by $\log\left(\frac{2}{\sigma_N^2}\frac{e^{-\frac{x^2}{\sigma_N^2}}}{x^2}\right)$ gives

$$h_{Y|X}^{(1)} = -\frac{4}{\sigma_S^2\sigma_N^2}\int_0^\infty \exp\left[-\left(\frac{\sigma_S^2+\sigma_N^2}{\sigma_S^2\sigma_N^2}\right)x^2\right]\log\left(\frac{2}{\sigma_N^2}\frac{e^{-\frac{x^2}{\sigma_N^2}}}{x^2}\right)\frac{\mathcal{I}^{(1)}(x)}{x}\,dx, \tag{D.18}$$

where the integral $\mathcal{I}^{(1)}(x)$ is defined as

$$\begin{aligned}\mathcal{I}^{(1)}(x) &\triangleq \int_0^\infty y^3 e^{-\frac{y^2}{\sigma_N^2}} I_2\left(\frac{2xy}{\sigma_N^2}\right) dy = \int_0^\infty y^3 e^{-p^2y^2} I_2(by)\,dy \\ &= \frac{b^2}{(2p^2)^3}\exp\left(\frac{b^2}{4p^2}\right) = \frac{\sigma_N^2}{2}x^2 e^{\frac{x^2}{\sigma_N^2}}. \end{aligned}\tag{D.19}$$

Thus, we have

$$\begin{aligned}h_{Y|X}^{(1)} &= -\frac{2}{\sigma_S^2}\int_0^\infty x\exp\left(-\frac{x^2}{\sigma_S^2}\right)\log\left(\frac{2}{\sigma_N^2}\frac{e^{-\frac{x^2}{\sigma_N^2}}}{x^2}\right)dx \\ &= h_{Y|X}^{(1,1)} + h_{Y|X}^{(1,2)} + h_{Y|X}^{(1,3)}, \end{aligned}\tag{D.20}$$

where

$$h_{Y|X}^{(1,1)} = \frac{2}{\sigma_S^2}\log\frac{\sigma_N^2}{2}\int_0^\infty x\exp\left(-\frac{x^2}{\sigma_S^2}\right)dx = \frac{2}{\sigma_S^2}\log\frac{\sigma_N^2}{2}\int_0^\infty xe^{-q^2x^2}\,dx$$
$$= \frac{2}{\sigma_S^2}\log\left(\frac{\sigma_N^2}{2}\right)\frac{1}{2q^2} = \log\frac{\sigma_N^2}{2}\,, \tag{D.21}$$

$$h_{Y|X}^{(1,2)} = \frac{4}{\sigma_S^2}\int_0^\infty x\exp\left(-\frac{x^2}{\sigma_S^2}\right)\log x\,dx$$
$$= \frac{4}{\sigma_S^2}\int_0^\infty xe^{-q^2x^2}\log x\,dx = \frac{1}{\sigma_S^2}\int_0^\infty e^{-q^2t}\log t\,dt$$
$$= -\frac{1}{\sigma_N^2}\,\frac{2\log q-\psi(1)}{q^2} = \log\sigma_S^2+\psi(1)\,, \tag{D.22}$$

$$h_{Y|X}^{(1,2)} = \frac{2}{\sigma_S^2\sigma_N^2}\int_0^\infty x^3\exp\left(-\frac{x^2}{\sigma_S^2}\right)dx = \frac{2}{\sigma_S^2\sigma_N^2}\int_0^\infty x^3e^{-q^2x^2}\,dx$$
$$= \frac{2}{\sigma_S^2\sigma_N^2}\,\frac{1}{2(q^2)^2} = \frac{\sigma_S^2}{\sigma_N^2}\,. \tag{D.23}$$

The final expression of the 1st conditional entropy contribution $h_{Y|X}^{(1)}$ reads

$$h_{Y|X}^{(1)} = \log\frac{\rho}{2}+\rho-\psi(1)\,. \tag{D.24}$$

The 2nd-term $h_{Y|X}^{(2)}$ is captured by $(3\log y)$ and given by

$$h_{Y|X}^{(2)} = -\frac{12}{\sigma_S^2\sigma_N^2}\int_0^\infty \exp\left[-\left(\frac{\sigma_S^2+\sigma_N^2}{\sigma_S^2\sigma_N^2}\right)x^2\right]\frac{\mathcal{I}^{(2)}(x)}{x}\,dx\,, \tag{D.25}$$

where the integral $\mathcal{I}^{(2)}(x)$ is given by

$$\begin{aligned}\mathcal{I}^{(2)}(x) &\triangleq \int_0^\infty y^3 e^{-\frac{y^2}{\sigma_N^2}} I_2\left(\frac{2xy}{\sigma_N^2}\right)\log y\,dy = \int_0^\infty y^3 e^{-p^2y^2}\, I_2(by)\log y\,dy \\ &= \sum_{k=0}^{\infty}\frac{1}{k!(k+2)!}\left(\frac{b}{2}\right)^{2k+2}\int_0^\infty y^{2k+5}e^{-p^2y^2}\log y\,dy \\ &= \frac{b^2}{2^4}\sum_{k=0}^{\infty}\frac{1}{k!(k+2)!}\left(\frac{b}{2}\right)^{2k}\int_0^\infty t^{k+2}e^{-p^2t}\log t\,dt \\ &= \frac{b^2}{2^4(p^2)^3}\sum_{k=0}^{\infty}\frac{\psi(k+3)-\log p^2}{k!}\left(\frac{b}{2p}\right)^{2k}. \end{aligned} \tag{D.26}$$

Next, we use the following decomposition of the digamma function

$$\psi(k+3) = \psi(k+1) + \frac{1}{k+1} + \frac{1}{k+2}, \tag{D.27}$$

and also we introduce the notation $u \triangleq \frac{b^2}{4p^2}$, then we have

$$\begin{aligned}\mathcal{I}^{(2)}(x) &= \frac{b^2}{16p^6}\left\{\sum_{k=0}^{\infty}\frac{\psi(k+1)}{k!}u^k - \log p^2\sum_{k=0}^{\infty}\frac{u^k}{k!} + \sum_{k=0}^{\infty}\frac{u^k}{(k+1)\,k!} + \sum_{k=0}^{\infty}\frac{u^k}{(k+2)\,k!}\right\} \\ &= \frac{b^2}{16p^6}\left\{\mathcal{I}^{(2,1)}(x) + \mathcal{I}^{(2,2)}(x) + \mathcal{I}^{(2,3)}(x)\right\}. \end{aligned} \tag{D.28}$$

Omitting the details of fairly cumbersome algebraic manipulations, we have the following analytical closed-form expressions for all these summations

$$\mathcal{I}^{(2,1)}(x) \triangleq \sum_{k=0}^{\infty}\frac{\psi(k+1)}{k!}u^k - \log p^2\sum_{k=0}^{\infty}\frac{u^k}{k!} = \left[\log\frac{u}{p^2} - \mathrm{Ei}(-u)\right]e^u, \tag{D.29}$$

$$\mathcal{I}^{(2,2)}(x) \triangleq \sum_{k=0}^{\infty}\frac{u^k}{(k+1)\,k!} = \frac{e^u-1}{u}, \tag{D.30}$$

$$\mathcal{I}^{(2,3)}(x) \triangleq \sum_{k=0}^{\infty}\frac{u^k}{(k+2)\,k!} = \frac{ue^u - e^u + 1}{u^2}. \tag{D.31}$$

Therefore, the integral $\mathcal{I}^{(2)}(x)$ is equal to

$$\begin{aligned}\mathcal{I}^{(2)}(x) &= \frac{b^2}{16p^2}\left[\log\left(\frac{b^2}{4p^2}\right) - \mathrm{Ei}\left(-\frac{b^2}{4p^2}\right)\right]\exp\left(\frac{b^2}{4p^2}\right) \\ &\quad + \frac{1}{2p^2p^4}\left[\left(b^2-2p^2\right)\exp\left(\frac{b^2}{4p^2}\right) + 2p^2 - \frac{b^2}{2}\right] \qquad \text{(D.32)} \\ &= \frac{\sigma_N^2}{4}x^2\left[2\log x - \mathrm{Ei}\left(-\frac{x^2}{\sigma_N^2}\right)\right]\exp\left(\frac{x^2}{\sigma_N^2}\right) \\ &\quad + \frac{\sigma_N^4}{4}\left[2\exp\left(\frac{x^2}{\sigma_N^2}\right)-1\right] - \frac{\sigma_n^6}{4x^2}\left[\exp\left(\frac{x^2}{\sigma_N^2}\right)-1\right]. \qquad \text{(D.33)}\end{aligned}$$

Thus, it is convenient to split the 2nd contribution $h_{Y|X}^{(2)}$ into the following three terms

$$h_{Y|X}^{(2)} = h_{Y|X}^{(2,1)} + h_{Y|X}^{(2,2)} + h_{Y|X}^{(2,3)}, \qquad \text{(D.34)}$$

where

$$\begin{aligned}h_{Y|X}^{(2,1)} &\triangleq -\frac{6}{\sigma_S^2}\int_0^\infty x\exp\left(-\frac{x^2}{\sigma_S^2}\right)\log x\,dx \\ &= -\frac{3}{2\sigma_S^2}\int_0^\infty \exp\left(-q^2t\right)\log t\,dt \\ &= \frac{3}{2\sigma_S^2}\left(\frac{2\log(q)-\psi(1)}{q^2}\right) = -\frac{3}{2}\left(\log\sigma_S^2+\psi(1)\right), \qquad \text{(D.35)}\end{aligned}$$

$$\begin{aligned}h_{Y|X}^{(2,2)} &\triangleq \frac{3}{\sigma_S^2}\int_0^\infty x\exp\left(-\frac{x^2}{\sigma_S^2}\right)\mathrm{Ei}\left(-\frac{x^2}{\sigma_N^2}\right)dx \\ &= -\frac{3}{2\sigma_S^2}\int_0^\infty \exp\left(-q^2t\right)E_1\left(p^2t\right)dt \\ &= -\frac{2}{2\sigma_S^2}\left[\frac{1}{q^2}\log\left(1+\frac{q^2}{p^2}\right)\right] = -\frac{3}{2}\log\left(1+\frac{\sigma_N^2}{\sigma_S^2}\right), \qquad \text{(D.36)}\end{aligned}$$

$$
\begin{aligned}
h_{Y|X}^{(2,3)} &\triangleq -\frac{3}{\sigma_S^2\sigma_N^2}\int_0^\infty \exp\left[-\left(\frac{\sigma_S^2+\sigma_N^2}{\sigma_S^2\sigma_N^2}\right)x^2\right] \\
&\quad\times\left\{\frac{\sigma_N^4}{x}\left[2\exp\left(\frac{x^2}{\sigma_N^2}\right)-1\right]-\frac{\sigma_N^6}{x^3}\left[\exp\left(\frac{x^2}{\sigma_N^2}\right)-1\right]\right\}dx \\
&= -\frac{3}{\sigma_S^2\sigma_N^2}\int_0^\infty\left\{\frac{\sigma_N^4}{x}\left[2-\exp\left(-\frac{x^2}{\sigma_N^2}\right)\right]\right. \\
&\quad\left.-\frac{\sigma_N^6}{x^3}\left[1-\exp\left(-\frac{x^2}{\sigma_N^2}\right)\right]\right\}\exp\left(-\frac{x^2}{\sigma_S^2}\right)dx \\
&= -\frac{3\sigma_N^2}{2\sigma_S^2}\left[\left(2+\frac{\sigma_N^2}{\sigma_S^2}\right)\log\left(1+\frac{\sigma_S^2}{\sigma_N^2}\right)-1\right]. \qquad \text{(D.37)}
\end{aligned}
$$

Finally, the 2nd contribution can be expressed as

$$
h_{Y|X}^{(2)} = -\frac{3}{2}\left[\left(1+\frac{1}{\rho}\right)^2\log(1+\rho)-\frac{1}{\rho}+\psi(1)\right]. \qquad \text{(D.38)}
$$

The 3rd-term $h_{Y|X}^{(3)}$ is captured by $\left(-\frac{y^2}{\sigma_N^2}\right)$ is given by the following integral

$$
h_{Y|X}^{(3)} = \frac{4}{\sigma_S^2\sigma_N^4}\int_0^\infty \exp\left[-\left(\frac{\sigma_S^2+\sigma_N^2}{\sigma_S^2\sigma_N^2}\right)x^2\right]\frac{\mathcal{I}^{(3)}(x)}{x}\,dx\,, \qquad \text{(D.39)}
$$

where the integral $\mathcal{I}^{(3)}(x)$ is given by

$$
\begin{aligned}
\mathcal{I}^{(3)}(x) &\triangleq \int_0^\infty y^5 e^{-\frac{y^2}{\sigma_N^2}} I_2\left(\frac{2xy}{\sigma_N^2}\right)dy \\
&= \int_0^\infty y^5 e^{-p^2y^2} I_2(by)\,dy \\
&= \frac{3b^2}{8p^8}\left(1+\frac{b^2}{12p^2}\right)e^{\frac{b^2}{4p^2}} = \frac{2\sigma_N^4}{2}x^2\left(1+\frac{x^2}{3\sigma_N^2}\right)e^{\frac{x^2}{\sigma_N^2}}. \qquad \text{(D.40)}
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
h_{Y|X}^{(3)} &= \frac{6}{\sigma_S^2}\int_0^\infty x\left(1+\frac{x^2}{3\sigma_N^2}\right)e^{-\frac{x^2}{\sigma_S^2}}dx \qquad \text{(D.41)} \\
&= h_{Y|X}^{(3,1)} + h_{Y|X}^{(3,2)}\,,
\end{aligned}
$$

where

$$h_{Y|X}^{(3,1)} = \frac{6}{\sigma_S^2}\int_0^\infty x e^{-\frac{x^2}{\sigma_S^2}}\,dx = \frac{6}{\sigma_S^2}\int_0^\infty x e^{-q^2x^2}\,dx$$
$$= \frac{6}{\sigma_S^2}\frac{1}{2q^2} = 3\,, \tag{D.42}$$

$$h_{Y|X}^{(3,2)} = \frac{2}{\sigma_S^2\sigma_N^2}\int_0^\infty x^3 e^{-\frac{x^2}{\sigma_S^2}}\,dx$$
$$= \frac{2}{\sigma_S^2\sigma_N^2}\int_0^\infty x^3 e^{-q^2x^2}\,dx = \frac{2}{\sigma_S^2\sigma_N^2}\frac{1}{2(q^2)^2} = \frac{\sigma_S^2}{\sigma_N^2}\,. \tag{D.43}$$

Thus, the final expression of $h_{Y|X}^{(3)}$ is

$$h_{Y|X}^{(3)} = \rho + 3\,. \tag{D.44}$$

The 4th-term $h_{Y|X}^{(4)}$ which is captured by $\log\left[I_2\left(\frac{2xy}{\sigma_N^2}\right)\right]$ arrives

$$h_{Y|X}^{(4)} = -\frac{4}{\sigma_S^2\sigma_N^2}\int_0^\infty\int_0^\infty \frac{y^3}{x} e^{-\frac{x^2}{\sigma_S^2}} e^{-\frac{x^2+y^2}{\sigma_N^2}} I_2\left(\frac{2xy}{\sigma_N^2}\right)\log\left[I_2\left(\frac{2xy}{\sigma_N^2}\right)\right]dxdy\,. \tag{D.45}$$

Let us make the change of variables $\xi \triangleq \frac{2xy}{\sigma_N^2}$, and $\eta \triangleq \frac{y^2}{\sigma_N^4}$. Then, Jacobian is given by $\mathcal{J} \triangleq \frac{\partial(x,y)}{\partial(\xi,\eta)} = \frac{\sigma_N^4}{4y^2}$. Substituting and changing the order of integration, the term $h_{Y|X}^{(4)}$ in the new variables (ξ,η) arrives at

$$h_{Y|X}^{(4)} = -\frac{2\sigma_N^2}{\sigma_S^2}\int_0^\infty I_2(\xi)\log\left[I_2(\xi)\right]\frac{\mathcal{I}^{(4)}(\xi)}{\xi}\,d\xi\,, \tag{D.46}$$

where

$$\mathcal{I}^{(4)}(\xi) \triangleq \int_0^\infty \eta\exp\left[-\eta - \frac{\xi^2}{4\eta}\left(1+\frac{\sigma_N^2}{\sigma_S^2}\right)\right]d\eta = \int_0^\infty \eta\exp\left[-\left(\frac{\xi}{4\eta}+\eta\right)\right]d\eta$$
$$= \frac{\xi}{2}K_2\left(\sqrt{\xi}\right) = \frac{1}{2}\left(1+\frac{\sigma_N^2}{\sigma_S^2}\right)\xi^2 K_2\left(\sqrt{1+\frac{\sigma_N^2}{\sigma_S^2}}\,\xi\right). \tag{D.47}$$

Thus, the 4th-term $h_{Y|X}^{(4)}$ finally reads

$$\begin{aligned} h_{Y|X}^{(4)} &= -\frac{\sigma_N^2}{\sigma_S^2}\left(1+\frac{\sigma_N^2}{\sigma_S^2}\right)\int_0^\infty \xi K_2\left(\sqrt{1+\frac{\sigma_N^2}{\sigma_S^2}}\,\xi\right) I_2(\xi)\log\left[I_2(\xi)\right]\,d\xi \\ &= -\frac{1}{\rho}\left(1+\frac{1}{\rho}\right)F_2(\rho)\,, \end{aligned} \tag{D.48}$$

where the function $F_2(\rho)$ is introduced as

$$F_2(\rho) \triangleq \int_0^\infty \xi K_2\left(\sqrt{1+\rho^{-1}}\,\xi\right) I_2(\xi)\log\left[I_2(\xi)\right]\,d\xi\,. \tag{D.49}$$

The proof is completed by combining the Eqs. (D.24), (D.38), (D.44), (D.46).

■

Appendix E

Proof of Lemmas 5.6.7 and 5.6.8

Proof of Lemma 5.6.7. The MI is invariant under a simultaneous linear re-scaling of the variables $x \to x/\sigma_N$ and $y \to y/\sigma_N$. For notation simplicity, and without loss of generality, throughout this proof we thus assume $\sigma_N^2 = 1$. Furthermore, we study the conditional entropy as a function of $\rho = \sigma_S^2$ and all the results will be given in [nats].

We express the conditional differential entropy as

$$h_{Y|X}(\rho) = -\int_0^\infty \int_0^\infty p_{X,Y}(x,y) \log p_{Y|X}(y|x)\, dx\, dy \tag{E.1}$$

$$\begin{aligned} &= -\log 2 - n\,\mathbb{E}[\log Y] + (n-1)\,\mathbb{E}[\log X] \\ &+ \mathbb{E}\left[X^2\right] + \mathbb{E}\left[Y^2\right] - \mathbb{E}[\log I_{n-1}(2XY)], \end{aligned} \tag{E.2}$$

where (E.2) follows from (5.39). In what follows, we will compute the 5 expectations in (E.2).

The third and fourth terms in (E.2) can be readily obtained using (5.44)

$$\mathbb{E}[\log X] = \frac{1}{2}(\log \rho + \psi(1)), \tag{E.3}$$

$$\mathbb{E}\left[X^2\right] = \rho. \tag{E.4}$$

In order to compute the second and fifth terms in (E.2), we first calculate the marginal output distribution as

$$p_Y(y) = \int_0^\infty p_{X,Y}(x,y)\, dx \tag{E.5}$$

$$= \frac{2y}{\rho \alpha^{n-2}} e^{-\frac{y^2}{\rho+1}} \left(1 - e^{-\alpha y^2} \sum_{k=0}^{n-2} \frac{(\alpha y^2)^k}{k!}\right), \tag{E.6}$$

where the joint distribution $p_{X,Y}(x,y)$ can be expressed using (5.39) and (5.44) as

$$\begin{aligned} p_{X,Y}(x,y) &= p_{Y|X}(x|y)\,p_X(x) \\ &= \frac{4}{\rho}\frac{y^n}{x^{n-2}}\exp\left(-\frac{x^2+\alpha y^2}{\alpha}\right) I_{n-1}(2xy), \end{aligned} \tag{E.7}$$

with

$$\alpha \triangleq \frac{\rho}{\rho+1} < 1, \tag{E.8}$$

and where (E.6) can be obtained using a symbolic integration software. Using (E.6), we obtain (using a symbolic integration software)

$$\mathbb{E}[\log Y] = \frac{1}{2}\left(\alpha\Phi(\alpha,1,n)+\psi(n)\right), \tag{E.9}$$

where $\psi(n)$ is the digamma function, $\Phi(\alpha,1,n)$ is given by Eq. (5.60). The second moment of the output distribution is obtained directly from the channel input-output relation (5.40), yielding

$$\mathbb{E}\left[Y^2\right] = \rho + n. \tag{E.10}$$

Substituting (E.3), (E.4), (E.9) and (E.10) into (E.2), we have

$$\begin{aligned} h_{Y|X}(\rho) = &-\log 2 - \frac{n}{2}\alpha\Phi(\alpha,1,n) - \frac{n}{2}\psi(n) \\ &+ \frac{n-1}{2}\left(\log\rho + \psi(1)\right) + 2\rho + n - h^{(6)}_{Y|X}(\rho), \end{aligned} \tag{E.11}$$

where

$$h^{(6)}_{Y|X}(\rho) \triangleq \int_0^\infty\int_0^\infty p_{X,Y}(x,y)\log\left[I_{n-1}(2xy)\right]dx\,dy \tag{E.12}$$

The last step is to compute the term $h^{(6)}_{Y|X}(\rho)$, which using (E.7) can be expressed as

$$h^{(6)}_{Y|X}(\rho) = \frac{4}{\rho}\int_0^\infty\int_0^\infty \frac{y^n}{x^{n-2}}\exp\left(-\frac{x^2+\alpha y^2}{\alpha}\right) I_{n-1}\left(2xy\right)\log\left[I_{n-1}(2xy)\right]dx\,dy. \tag{E.13}$$

We then make the change of variables $\xi = 2xy,\ \eta = y^2$, with the Jacobian $\partial(x,y)/\partial(\xi,\eta) = (4y^2)^{-1}$, yielding

$$h^{(6)}_{Y|X}(\rho) = \frac{2^{n-2}}{\rho} \int_0^\infty I_{n-1}(\xi) \log\left[I_{n-1}(\xi)\right] \int_0^\infty \left(\frac{\eta}{\xi}\right)^{n-2} \exp\left(-\frac{\xi^2}{4\eta\alpha} - \eta\right) d\eta\, d\xi. \tag{E.14}$$

The integration over η can be performed analytically, yielding

$$\int_0^\infty \left(\frac{\eta}{\xi}\right)^{n-2} \exp\left(-\frac{\xi^2}{4\eta\alpha} - \eta\right) d\eta = 2^{2-n} \alpha^{(1-n)/2}\, \xi\, K_{n-1}\left(\frac{\xi}{\alpha^{1/2}}\right), \tag{E.15}$$

where $K_n(x)$ is the modified Bessel function of the second kind of order n. Using (E.15) in (E.14) gives

$$h^{(6)}_{Y|X}(\rho) = \frac{\alpha^{(1-n)/2}}{\rho} \int_0^\infty \xi K_{n-1}\left(\frac{\xi}{\alpha^{1/2}}\right) I_{n-1}(\xi) \log\left[I_{n-1}(\xi)\right] d\xi \tag{E.16}$$

$$= \frac{\alpha^{(1-n)/2}}{\rho} F_n(\rho). \tag{E.17}$$

The proof is completed by using (E.17) in (E.11), the definition of α in (E.8), and by returning to logarithm base 2. ■

Proof of Lemma 5.6.8. From (E.6), it follows that the output entropy can then be expressed as[1]

$$h_Y(\rho) = \log\left(\frac{\rho\alpha^{n-2}}{2}\right) - \mathbb{E}\left[\log Y\right] + \frac{1}{\rho+1}\,\mathbb{E}\left[Y^2\right] + h_Y^{(4)}(\rho), \tag{E.18}$$

where α is given by (E.8),

$$h_Y^{(4)}(\rho) \triangleq \mathbb{E}\left[g_Y^{(4)}(Y)\right] \tag{E.19}$$

$$= \int_0^\infty p_X(x) \int_0^\infty p_{Y|X}(y|x)\, g_Y^{(4)}(y)\, dy\, dx \tag{E.20}$$

$$= \int_0^\infty p_Y(y)\, g_Y^{(4)}(y)\, dy, \tag{E.21}$$

where $p_Y(y)$ is given by (E.6) and

$$g_Y^{(4)}(y) \triangleq -\log f(\alpha y^2) \tag{E.22}$$

$$f(z) \triangleq 1 - e^{-z} \sum_{k=0}^{n-2} \frac{(z)^k}{k!}. \tag{E.23}$$

Notice that from its definition it follows that the function $f(z)$ is confined to the interval $0 \leq f(z) \leq 1$. We shall now prove that $h_Y^{(4)}(\rho)$ decays as $O\left[\rho^{-1}\right]$ or faster when $\rho \to \infty$. Indeed, one has

$$h_Y^{(4)}(\rho) = -\int_0^\infty \frac{2y}{\rho\alpha^{n-2}}\, e^{-\frac{y^2}{\rho+1}}\, f\left(\alpha y^2\right) \log f\left(\alpha y^2\right)\, dy \tag{E.24}$$

$$= -\frac{1}{\rho\alpha^{n-1}} \int_0^\infty e^{-z/\rho}\, f(z) \log f(z)\, dz. \tag{E.25}$$

Next, one notices that $h_Y^{(4)}(\rho)$ is positive and can be upper-bounded as follows

$$h_Y^{(4)}(\rho) \leq \frac{1}{\rho\alpha^{n-1}} \int_0^\infty \left(-f(z)\log f(z)\right) dz \tag{E.26}$$

$$\triangleq \frac{C}{\rho\alpha^{n-1}}. \tag{E.27}$$

It is therefore only left to prove that the integral converges, *i.e.,* that the con-

[1]Similarly to the previous Lemma, the results in this proof are in given in [nats].

stant C is finite. This can be done as follows:

$$\begin{aligned} C &= \int_0^\infty (-f(z)\log f(z))\, dz \\ &\leq \int_0^\infty (1-f(z))\, dz \\ &= \int_0^\infty e^{-z} \sum_{k=0}^{n-2} \frac{z^k}{k!}\, dz \\ &= n-1 \\ &< \infty, \end{aligned}$$

where in the second line we have used an inequality $-x\ln x \leq (1-x)$, $x \in (0,1]$. Therefore, asymptotically $h_Y^{(4)}(\rho)$ decays not slower than $1/\rho$.

The asymptotic expression for the output entropy can be written by combining (E.26), (E.9), (E.10) and (E.18), which yields

$$h_Y(\rho) = \frac{1}{2}\log\rho + 1 - \frac{\psi(1)}{2} - \log 2 + O\left[\rho^{-1}\right]. \tag{E.28}$$

The proof is completed by returning to logarithm base 2. It should also be noted that this asymptote is independent on the order n. ■

Acronyms

AIR achievable information rate

ASE amplified spontaneous emission

ASK amplitude-shift keying

AWGN additive white Gaussian noise

BER bit error ratio

BW bandwidth

DBP digital back propagation

DP dual polarisation

DRA distributed Raman amplifier

EDC electrical dispersion compensation

EDFA erbium-doped fiber amplifier

EM electro-magnetic

FF Full-field

FT Fourier transform

FWM four-wave mixing

GH Gordon-Haus

GN Gaussian noise

GVD group velocity dispersion

IDRA ideal distributed Raman amplifier

IID independent identically distributed

ISI inter-symbol interference

IQ in-phase and quadrature

ME Manakov equation

MI mutual information

MZM Mach-Zehnder modulator

NF noise figure

NFT nonlinear Fourier transform

NL nonlinear

NLC nonlinearity compensation

NLI nonlinear interference

NLSE nonlinear Schrödinger equation

OFDM orthogonal frequency-division multiplexing

OOK on-off keying

OSNR optical signal-to-noise ratio

PDF probability density function

PMD polarisation mode dispersion

PMF probability mass function

PS probabilistic shaping

PSD power spectral density

QAM quadrature amplitude modulation

QPSK quadrature phase-shift keying

RRC root-raised-cosine

RV random variable

SMF single mode fibre

SNI signal-noise interaction

SNR signal-to-noise ratio

SPM self-phase modulation

SSFT split-step Fourier transform

SSI signal-signal interaction

SSMF standard single-mode fibre

SVE slowly-varying envelope

SVEA slowly varying envelope approximation

WDM wavelength-division multiplexing

XPM cross-phase modulation

Bibliography

[1] TeleGeography. Submarine cable map. *[Online] Available: https://www.submarinecablemap.com/.*

[2] Cisco Systems Inc. White paper: Cisco visual networking index: Forecast and methodology, 2016-2021. *[Online] Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html*, 2017.

[3] R. N. Hall, G. E. Fenner, J. D. Kingsley, T. J. Soltys, and R. O. Carlson. Coherent light emission from GaAs junctions. *Physical Review Letters*, 9(9):366, 1962.

[4] N. Holonyak Jr. and S. F. Bevacqua. Coherent (visible) light emission from $Ga(As_{1-x}P_x)$ junctions. *Applied Physics Letters*, 1(4):82–83, 1962.

[5] I. Hayashi, M. B. Panish, P. W. Foy, and S. Sumski. Junction lasers which operate continuously at room temperature. *Applied Physics Letters*, 17(3):109–111, 1970.

[6] K. C. Kao and G. A. Hockham. Dielectric-fibre surface waveguides for optical frequencies. In *Proceedings of the Institution of Electrical Engineers*, volume 113, pages 1151–1158. IET, 1966.

[7] T. Miya, Y. Terunuma, T. Hosaka, and T. Miyashita. Ultimate low-loss single-mode fibre at 1.55 μm. *Electron. Lett.*, 15(4):106–108, 1979.

[8] S. Makovejs, C. C. Roberts, F. Palacios, H. B. Matthews, D. A. Lewis, D. T. Smith, P. G. Diehl, J. J. Johnson, J. D. Patterson, C. R. Towery, et al. Record-low (0.1460 dB/km) attenuation ultra-large A_{eff} optical fiber for submarine applications. In *Optical Fiber Communications Conference and Exhibition (OFC), 2015*, pages 1–3. IEEE, 2015.

[9] Y. Tamura, H. Sakuma, K. Morita, M. Suzuki, Y. Yamamoto, K. Shimada, Y. Honma, K. Sohma, T. Fujii, and T. Hasegawa. The first 0.14-dB/km loss optical fiber and its impact on submarine transmission. *Journal of Lightwave Technology*, 36(1):44–49, 2018.

[10] S. B. Poole, D. N. Payne, R. J. Mears, M. E. Fermann, and R. I. Laming. Fabrication and characterization of low-loss optical fibers containing rare-earth ions. *Journal of Lightwave Technol*, 4(7):870–876, 1986.

[11] D. Payne, R. J. Mears, L. Reeky, and I. M. Jauncy. Low noise erbium doped fiber amplifier operating at 1.54 μm. *Electronic Letters*, 23(19):1026âĂŞ–1027, 1987.

[12] E. Desurvire, J. R. Simpson, and P. C. Becker. High-gain erbium-doped traveling-wave fiber amplifier. *Optics Letters*, 12(11):888–890, 1987.

[13] R. H. Stolen and E. P. Ippen. Raman gain in glass optical waveguides. *Applied Physics Letters*, 22(6):276–278, 1973.

[14] Y. Aoki. Properties of fiber Raman amplifiers and their applicability to digital optical communication systems. *Journal of Lightwave Technology*, 6(7):1225–1239, 1988.

[15] J. Chesnoy. *Undersea fiber communication systems*. Academic press, 2015.

[16] M. Cvijetic and I. Djordjevic. *Advanced optical communication systems and networks*. Artech House, Norwood, MA, 2013.

[17] R. H. Stolen and A. Ashkin. Optical Kerr effect in glass waveguide. *Applied Physics Letters*, 22(6):294–296, 1973.

[18] D. J. Richardson. Filling the light pipe. *Science*, 330(6002):327–328, 2010.

[19] A. D. Ellis, J. Zhao, and D. Cotter. Approaching the non-linear Shannon limit. *Journal of Lightwave Technology*, 28(4):423–433, 2010.

[20] A. Mecozzi and R.-J. Essiambre. Nonlinear Shannon limit in pseudo-linear coherent systems. *Journal of Lightwave Technology*, 30(12):2011–2024, 2012.

[21] E. Agrell, M. Karlsson, A. R. Chraplyvy, D. J. Richardson, P. M. Krummrich, P. Winzer, K. Roberts, J. K. Fischer, S. J. Savory, B. J. Eggleton, et al. Roadmap of optical communications. *Journal of Optics*, 18(6):063002, 2016.

[22] A. Chraplyvy. Plenary paper: The coming capacity crunch. In *Optical Communication, 2009. ECOC'09. 35th European Conference on*, pages 1–1. IEEE, 2009.

[23] E. Agrell, A. Alvarado, and F. R. Kschischang. Implications of information theory in optical fibre communications. *Phil. Trans. R. Soc. A*, 374(20140438), 2016.

[24] E. Forestieri and M. Secondini. The nonlinear fiber-optic channel: Modeling and achievable information rate. In *Progress In Electromagnetics Research Symposium (PIERS), Prague, Czech Republic*, 2015.

[25] M. Secondini and E. Forestieri. Scope and limitations of the nonlinear Shannon limit. *Journal of Lightwave Technology*, 35(4):893–902, 2017.

[26] N. A. Shevchenko, T. Xu, D. Lavery, G. Liga, D. J. Ives, R. I. Killey, and P. Bayvel. Modeling of nonlinearity-compensated optical communication systems considering second-order signal-noise interactions. *Optics Letters*, 42(17):3351–3354, 2017.

[27] N. A. Shevchenko, T. Xu, D. Semrau, G. Saavedra, G. Liga, M. Paskov, L. Galdino, A. Alvarado, R. I. Killey, and P. Bayvel. Achievable information rates estimation for 100-nm Raman-amplified optical transmission system. In *ECOC 2016; 42nd European Conference on Optical Communication; Proceedings of*, pages 1–3. VDE, 2016.

[28] D. Semrau, T. Xu, N. A. Shevchenko, M. Paskov, A. Alvarado, R. I. Killey, and P. Bayvel. Achievable information rates estimates in optically amplified transmission systems using nonlinearity compensation and probabilistic shaping. *Optics Letters*, 42(1):121–124, 2017.

[29] P. Bayvel, R. Maher, T. Xu, G. Liga, N. A. Shevchenko, D. Lavery, A. Alvarado, and R. I. Killey. Maximizing the optical network capacity. *Phil. Trans. R. Soc. A*, 374(2062):20140440, 2016.

[30] N. A. Shevchenko, J. E. Prilepsky, S. A. Derevyanko, A. Alvarado, P. Bayvel, and S. K. Turitsyn. A lower bound on the per soliton ca-

pacity of the nonlinear optical fibre channel. In *Information Theory Workshop-Fall (ITW), 2015 IEEE*, pages 104–108. IEEE, 2015.

[31] N. A. Shevchenko, S. A. Derevyanko, J. E. Prilepsky, A. Alvarado, P. Bayvel, and S. K. Turitsyn. Capacity lower bounds of the noncentral chi-channel with applications to soliton amplitude modulation. *IEEE Transactions on Communications*, 2018.

[32] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[33] R. E. Blahut. *Principles and practice of information theory*. Addison-Wesley Longman Publishing Co., Inc., 1987.

[34] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[35] S. Verdú and Han T. S. A general formula for channel capacity. *IEEE Transactions on Information Theory*, 40(4):1147–1157, 1994.

[36] W. Hirt and J. L. Massey. Capacity of the discrete-time Gaussian channel with intersymbol interference. *IEEE Transactions on Information Theory*, 34(3):38–38, 1988.

[37] A. Y. Khinchin. *Mathematical foundations of information theory*. Dover, New York, 2013.

[38] C. Kittel, P. McEuen, and P. McEuen. *Introduction to solid state physics*, volume 8. Wiley, New York, 1996.

[39] D. Marcuse. *Light transmission optics*. Van Nostrand Reinhold, New York, 1982.

[40] I. H. Malitson. Interspecimen comparison of the refractive index of fused silica. *Journal of the Optical Society of America*, 55(10):1205–1209, 1965.

[41] J. M. Dziedzic, R. H. Stolen, and A. Ashkin. Optical Kerr effect in long fibers. *Applied Optics*, 20(8):1403–1406, 1981.

[42] N. G. R. Broderick, T. M. Monro, P. J. Bennett, and D. J. Richardson. Nonlinearity in holey optical fibers: measurement and future opportunities. *Optics Letters*, 24(20):1395–1397, 1999.

[43] L. D. Landau and E. M. Lifshitz. *Electrodynamics of continuous media*, volume 8. Pergamon, New York, 1984.

[44] A. Shabat and V. Zakharov. Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media. *Sov. Phys. - JETP*, 34(1):62–69, 1972.

[45] G. B. Agrawal. *Nonlinear fiber optics*. Academic Press, London, 2001.

[46] A. Hasegawa and F. Tappert. Transmission of stationary nonlinear optical pulses in dispersive dielectric fibers. i. anomalous dispersion. *Applied Physics Letters*, 23(3):142–144, 1973.

[47] A. Hasegawa and F. Tappert. Transmission of stationary nonlinear optical pulses in dispersive dielectric fibers. ii. normal dispersion. *Applied Physics Letters*, 23(4):171–172, 1973.

[48] Antonio Mecozzi. Limits to long-haul coherent transmission set by the Kerr nonlinearity and noise of the in-line amplifiers. *Journal of Lightwave Technology*, 12(11):1993–2000, 1994.

[49] C. R. Menyuk. Nonlinear pulse propagation in birefringent optical fibers. *IEEE Journal of Quantum Electronics*, 23(2):174–176, 1987.

[50] C. R. Menyuk. Pulse propagation in an elliptically birefringent Kerr medium. *IEEE Journal of Quantum Electronics*, 25(12):2674–2682, 1989.

[51] D. Marcuse, C. R. Manyuk, and P. K. A. Wai. Application of the Manakov-PMD equation to studies of signal propagation in optical fibers with randomly varying birefringence. *Journal of Lightwave Technology*, 15(9):1735–1746, 1997.

[52] P. K. A. Wai, C. R. Menyuk, and H. H. Chen. Stability of solitons in randomly varying birefringent fibers. *Optics Letters*, 16(16):1231–1233, 1991.

[53] P. K. A. Wai, C. R. Menyuk, and H. H. Chen. Effects of randomly varying birefringence on soliton interactions in optical fibers. *Optics Letters*, 16(22):1735–1737, 1991.

[54] M. Ohashi, K. Shiraki, and K. Tajima. Optical loss property of silica-based single-mode fibers. *Journal of Lightwave Technology*, 10(5):539–543, 1992.

[55] M. E. Lines. Scattering losses in optic fiber materials. i. a new parameterization. *Journal of Applied Physics*, 55(11):4052–4057, 1984.

[56] M. Born and E. Wolf. *Principles of Optics*. Cambridge University Press, 2013.

[57] R. H. Hardin. Applications of the split-step Fourier method to the numerical solution of nonlinear and variable coefficient wave equations. *SIAM Review (Chronicles)*, 15(1):423–428, 1973.

[58] G. H. Weiss and A. A. Maradudin. The Baker-Hausdorff formula and a problem in crystal physics. *Journal of Mathematical Physics*, 3(4):771–777, 1962.

[59] J. A. Fleck, J. R. Morris, and M. D. Feit. Time-dependent propagation of high energy laser beams through the atmosphere. *Applied Physics*, 10(2):129–160, 1976.

[60] M. Seimetz. *High-order modulation for optical fiber transmission*, volume 143. Springer, 2009.

[61] M. Secondini, E. Forestieri, and G. Prati. Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps. *Journal of Lightwave Technology*, 31(23):3839–3852, 2013.

[62] A. Alvarado, E. Agrell, D. Lavery, R. Maher, and P. Bayvel. Replacing the soft-decision FEC limit paradigm in the design of optical communication systems. *Journal of Lightwave Technology*, 33(20):4338–4352, 2015.

[63] A. Alvarado, T. Fehenberger, B. Chen, and F. M. J. Willems. Achievable information rates for fiber optics: applications and computations. *Journal of Lightwave Technology*, 36(2):424–439, 2018.

[64] T. Fehenberger, A. Alvarado, P. Bayvel, and N. Hanik. On achievable rates for long-haul fiber-optic communications. *Optics Express*, 23(7):9183–9191, 2015.

[65] Keang-Po Ho and Han-Wei Cuei. Generation of arbitrary quadrature signals using one dual-drive modulator. *Journal of Lightwave Technology*, 23(2):764, 2005.

[66] E. Ip and J. M. Kahn. Compensation of dispersion and nonlinear impairments using digital backpropagation. *Journal of Lightwave Technology*, 26(20):3416–3425, 2008.

[67] S. J. Savory, G. Gavioli, R. I. Killey, and P. Bayvel. Electronic compensation of chromatic dispersion using a digital coherent receiver. *Optics Express*, 15(5):2120–2126, 2007.

[68] T. Xu, G. Jacobsen, S. Popov, J. Li, E. Vanin, K. Wang, A. T Friberg, and Y. Zhang. Chromatic dispersion compensation in coherent transmission system using digital filters. *Optics Express*, 18(15):16243–16257, 2010.

[69] G. Grimmett and D. Stirzaker. *Probability and random processes*. Oxford university press, 2001.

[70] G. D. Forney, R. Gallager, G. Lang, F. Longstaff, and S. Qureshi. Efficient modulation for band-limited channels. *IEEE Journal on Selected Areas in Communications*, 2(5):632–647, 1984.

[71] U. Wachsmann, R. F. H. Fischer, and J. B. Huber. Multilevel codes: Theoretical concepts and practical design rules. *IEEE Transactions on Information Theory*, 45(5):1361–1391, 1999.

[72] T. Fehenberger, A. Alvarado, G. Böcherer, and N. Hanik. On probabilistic shaping of quadrature amplitude modulation for the nonlinear fiber channel. *Journal of Lightwave Technology*, 34(21):5063–5073, 2016.

[73] D.-M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavcic, and W. Zeng. Simulation-based computation of information rates for channels with memory. *IEEE Transactions on Information Theory*, 52(8):3498–3508, 2006.

[74] N. Merhav, G. Kaplan, A. Lapidoth, and S. S. Shitz. On information rates for mismatched decoders. *IEEE Transactions on Information Theory*, 40(6):1953–1967, 1994.

[75] G. P. Agrawal. *Lightwave technology: telecommunication systems*. John Wiley & Sons, 2005.

[76] H. A. Haus. Quantum noise in a solitonlike repeater system. *Journal of the Optical Society of America B*, 8(5):1122–1126, 1991.

[77] E. Desurvire, D. Bayart, B. Desthieux, and S. Bigo. *Erbium-doped fiber amplifiers: Device and System Developments*. Wiley, New York, 2002.

[78] H. Kogelnik and A. Yariv. Considerations of noise and schemes for its reduction in laser amplifiers. *Proceedings of the IEEE*, 52(2):165–172, 1964.

[79] A. Yariv. Signal-to-noise considerations in fiber links with periodic or distributed optical amplification. *Optics Letters*, 15(19):1064–1066, 1990.

[80] P. Bayvel and P. M. Radmore. Solutions of the sbs equations in single mode optical fibres and implications for fibre transmission systems. *Electronics Letters*, 26(7):434–436, 1990.

[81] S. R. Chinn. Analysis of counter-pumped small-signal fibre Raman amplifiers. *Electronics Letters*, 33(7):607–608, 1997.

[82] J. Bromage. Raman amplification for fiber communications systems. *Journal of Lightwave Technology*, 22(1):79, 2004.

[83] E. Iannone, F. Matera, A. Mecozzi, and M. Settembre. *Nonlinear optical communication networks*. Wiley, New York, 1998.

[84] A. Mecozzi, C. B. Clausen, and M. Shtaif. System impact of intra-channel nonlinear effects in highly dispersed optical pulse transmission. *IEEE Photonics Technology Letters*, 12(12):1633–1635, 2000.

[85] E. E. Narimanov and P. Mitra. The channel capacity of a fiber optics communication system: Perturbation theory. *Journal of Lightwave Technology*, 20(3):530, 2002.

[86] A. Vannucci, P. Serena, and A. Bononi. The RP method: A new tool for the iterative solution of the nonlinear Schrödinger equation. *Journal of Lightwave Technology*, 20(7):1102, 2002.

[87] G. Bosco, P. Poggiolini, A. Carena, V. Curri, and F. Forghieri. Analytical results on channel capacity in uncompensated optical links with coherent detection. *Optics Express*, 19(26):B440–B451, 2011.

[88] T. Hasegawa, Y. Yamamoto, and M. Hirano. Optimal fiber design for large capacity long haul coherent transmission. *Optics Express*, 25(2):706–712, 2017.

[89] A. Splett, C. Kurtzke, and K. Petermann. Ultimate transmission capacity of amplified optical fiber communication systems taking into account fiber nonlinearities. In *ECOC 1993; The European Conference on Optical Communication; Proceedings of*, 1993.

[90] P. P. Mitra and J. B. Stark. Nonlinear limits to the information capacity of optical fibre communications. *Nature*, 411(6841):1027, 2001.

[91] J. B. Stark, P. Mitra, and A. Sengupta. Information capacity of nonlinear wavelength division multiplexing fiber optic transmission line. *Optical Fiber Technology*, 7(4):275–288, 2001.

[92] L. G. L. Wegener, B. M. L. Povinelli, A. G. Green, P. P. Mitra, J. B. Stark, and P. B. Littlewood. The effect of propagation nonlinearities on the information capacity of wdm optical fiber systems: Cross-phase modulation and four-wave mixing. *Physica D: Nonlinear Phenomena*, 189(1-2):81–99, 2004.

[93] J. Tang. The channel capacity of a multispan dwdm system employing dispersive nonlinear optical fibers and an ideal coherent optical receiver. *Journal of Lightwave Technology*, 20(7):1095, 2002.

[94] H. Louchet, A. Hodzic, and K. Petermann. Analytical model for the performance evaluation of DWDM transmission systems. *IEEE Photonics Technology Letters*, 15(9):1219–1221, 2003.

[95] X. Chen and W. Shieh. Closed-form expressions for nonlinear transmission performance of densely spaced coherent optical OFDM systems. *Optics Express*, 18(18):19039–19054, 2010.

[96] P. Poggiolini, A. Carena, V. Curri, G. Bosco, and F. Forghieri. Analytical modeling of nonlinear propagation in uncompensated optical transmission links. *IEEE Photonics Technology Letters*, 23(11):742–744, 2011.

[97] A. Carena, V. Curri, G. Bosco, and F Poggiolini, P.and Forghieri. Modeling of the impact of nonlinear propagation effects in uncompensated optical coherent transmission links. *Journal of Lightwave Technology*, 30(10):1524–1539, 2012.

[98] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri. The GN-model of fiber non-linear propagation and its applications. *Journal of Lightwave Technology*, 32(4):694–721, 2014.

[99] P. Johannisson and M. Karlsson. Perturbation analysis of nonlinear propagation in a strongly dispersive optical communication system. *Journal of Lightwave Technology*, 31(8):1273–1282, 2013.

[100] P. Johannisson and E. Agrell. Modeling of nonlinear signal distortion in fiber-optic networks. *Journal of Lightwave Technology*, 32(23):3942–3950, 2014.

[101] J. G. Proakis and M. Salehi. *Digital communication*. McGraw-Hill, Boston, MA, USA, 2008.

[102] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif. Properties of nonlinear noise in long, dispersion-uncompensated fiber links. *Optics Express*, 21(22):25685–25699, 2013.

[103] M. Secondini and E. Forestieri. Analytical fiber-optic channel model in the presence of cross-phase modulation. *IEEE Photonics Technology Letters*, 24(22):2016–2019, 2012.

[104] A. Bononi and P. S. N. Rossi. Performance dependence on channel baudrate of coherent single-carrier WDM systems. In *Optical Communication (ECOC 2013), 39th European Conference and Exhibition on*, pages 1–3. IET, 2013.

[105] P. Poggiolini, A. Nespola, Y. Jiang, G. Bosco, A. Carena, L. Bertignono, S. M. Bilal, S. Abrate, and F. Forghieri. Analytical and experimental results on system maximum reach increase through symbol rate optimization. *Journal of Lightwave Technology*, 34(8):1872–1885, 2016.

[106] E. Agrell, A. Alvarado, G. Durisi, and M. Karlsson. Capacity of a nonlinear optical channel with finite memory. *Journal of Lightwave Technology*, 32(16):2862–2876, 2014.

[107] R. Dar, M. Feder, Antonio Mecozzi, and M. Shtaif. Accumulation of nonlinear interference noise in fiber-optic systems. *Optics Express*, 22(12):14199–14211, 2014.

[108] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif. Inter-channel nonlinear interference noise in WDM systems: modeling and mitigation. *Journal of Lightwave Technology*, 33(5):1044–1053, 2015.

[109] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif. Pulse collision picture of inter-channel nonlinear interference in fiber-optic communications. *Journal of Lightwave Technology*, 34(2):593–607, 2016.

[110] R. E. Caflisch. Monte Carlo and quasi-Monte Carlo methods. *Acta numerica*, 7:1–49, 1998.

[111] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri. EGN model of non-linear fiber propagation. *Optics Express*, 22(13):16335–16362, 2014.

[112] P. Serena and A. Bononi. An alternative approach to the gaussian noise model and its system implications. *Journal of Lightwave Technology*, 31(22):3489–3499, 2013.

[113] P. Serena and A. Bononi. A time-domain extended gaussian noise model. *Journal of Lightwave Technology*, 33(7):1459–1472, 2015.

[114] W. Zeiler, F. Di Pasquale, P. Bayvel, and J. E. Midwinter. Modeling of four-wave mixing and gain peaking in amplified WDM optical communication systems and networks. *Journal of Lightwave Technology*, 14(9):1933–1942, 1996.

[115] W. S. Pelouch. Raman amplification: An enabling technology for long-haul coherent transmission systems. *Journal of Lightwave Technology*, 34(1):6–19, 2016.

[116] S. Popov, S. Sergeyev, and A. T. Friberg. The impact of pump polarization on the polarization dependence of the Raman gain due to the breaking of a fibre's circular symmetry. *Journal of Optics A: Pure and Applied Optics*, 6(3):S72, 2004.

[117] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products*. Academic press, New York, 2007.

[118] S. J. Savory. Approximations for the nonlinear self-channel interference of channels with rectangular spectra. *IEEE Photonics Technology Letters*, 25(10):961–964, 2013.

[119] G. Gao, X. Chen, and W. Shieh. Influence of PMD on fiber nonlinearity compensation using digital back propagation. *Optics Express*, 20(13):14406–14418, 2012.

[120] C. B. Czegledi, G. Liga, D. Lavery, M. Karlsson, E. Agrell, S. J. Savory, and P. Bayvel. Digital backpropagation accounting for polarization-mode dispersion. *Optics Express*, 25(3):1903–1915, 2017.

[121] M. A. Z. Al-Khateeb, M. McCarthy, C. Sánchez, and A. Ellis. Effect of second order signal-noise interactions in nonlinearity compensated optical transmission systems. *Optics Letters*, 41(8):1849–1852, 2016.

[122] D. Rafique and A. D. Ellis. Impact of signal-ASE four-wave mixing on the effectiveness of digital back-propagation in 112 Gb/s PM-QPSK systems. *Optics Express*, 19(4):3449–3454, 2011.

[123] T. Xu, N. A. Shevchenko, D. Lavery, D. Semrau, G. Liga, A. Alvarado, R. I. Killey, and P. Bayvel. Modulation format dependence of digital nonlinearity compensation performance in optical fibre communication systems. *Optics Express*, 25(4):3311–3326, 2017.

[124] A. D. Ellis, M. E. McCarthy, M. A. Z. Al-Khateeb, and S. Sygletos. Capacity limits of systems employing multiple optical phase conjugators. *Optics Express*, 23(16):20381–20393, 2015.

[125] D. Lavery, D. Ives, G. Liga, A. Alvarado, S. J. Savory, and P. Bayvel. The benefit of split nonlinearity compensation for single channel optical fiber communications. In *Photonics Conference (IPC), 2016 IEEE*, pages 799–802. IEEE, 2016.

[126] T. Tanimura, M. Nölle, J. K. Fischer, and C. Schubert. Analytical results on back propagation nonlinear compensator with coherent detection. *Optics Express*, 20(27):28779–28785, 2012.

[127] E. Agrell. Capacity bounds in optical communications. In *2017 European Conference on Optical Communication (ECOC)*, pages 1–3. IEEE, 2017.

[128] R. Dar and P. . Winzer. Nonlinear interference mitigation: Methods and potential gain. *Journal of Lightwave Technology*, 35(4):903–930, 2017.

[129] G. Bosco, A. Carena, V. Curri, R. Gaudino, P. Poggiolini, and S. Benedetto. Suppression of spurious tones induced by the split-step method in fiber systems simulation. *IEEE Photonics Technology Letters*, 12(5):489–491, 2000.

[130] R. Kudo, T. Kobayashi, K. Ishihara, Y. Takatori, A. Sano, and Y. Miyamoto. Coherent optical single carrier transmission using overlap frequency domain equalization for long-haul optical systems. *Journal of Lightwave Technology*, 27(16):3721–3728, 2009.

[131] R. Maher, T. Xu, L. Galdino, M. Sato, A. Alvarado, K. Shi, S. J. Savory, B. C. Thomsen, R. I. Killey, and P. Bayvel. Spectrally shaped DP-16QAM super-channel transmission with multi-channel digital backpropagation. *Scientific Reports*, 5:8214, 2015.

[132] T. Xu, B. Karanov, N. A. Shevchenko, D. Lavery, G. Liga, R. I. Killey, and P. Bayvel. Digital nonlinearity compensation in high-capacity optical communication systems considering signal spectral broadening effect. *Scientific Reports*, 7(1):12986, 2017.

[133] R.-J. Essiambre and R. W. Tkach. Capacity trends and limits of optical communication networks. *Proceedings of the IEEE*, 100(5):1035–1055, 2012.

[134] G. Liga, A. Alvarado, E. Agrell, and P. Bayvel. Information rates of next-generation long-haul optical fiber systems using coded modulation. *Journal of Lightwave Technology*, 35(1):113–123, 2017.

[135] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel. Capacity limits of optical fiber networks. *Journal of Lightwave Technology*, 28(4):662–701, 2010.

[136] F. R. Kschischang and S. Pasupathy. Optimal nonuniform signaling for gaussian channels. *IEEE Transactions on Information Theory*, 39(3):913–929, 1993.

[137] V. Curri, A. Carena, P. Poggiolini, G. Bosco, and F. Forghieri. Extension and validation of the GN model for non-linear interference to uncompensated links using raman amplification. *Optics Express*, 21(3):3308–3317, 2013.

[138] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri. A simple and effective closed-form GN model correction formula accounting for signal non-Gaussian distribution. *Journal of Lightwave Technology*, 33(2):459–473, 2015.

[139] J.-X. Cai, Y. Sun, H. Zhang, H. G. Batshon, M. V. Mazurczyk, O. V. Sinkin, D. G. Foursa, and A. Pilipetskii. 49.3 Tb/s transmission over 9100 km using C+L EDFA and 54 Tb/s transmission over 9150 km using hybrid-Raman EDFA. *Journal of Lightwave Technology*, 33(13):2724–2734, 2015.

[140] G. Jacobsen, M. Lidn, T. Xu, S. Popov, A. T. Friberg, and Y. Zhang. Influence of pre-and post-compensation of chromatic dispersion on equalization enhanced phase noise in coherent multilevel systems. *Journal of Optical Communications*, 32(4):257–261, 2011.

[141] G. Jacobsen, T. Xu, S. Popov, J. Li, A. T. Friberg, and Y. Zhang. EEPN and CD study for coherent optical nPSK and nQAM systems with RF pilot based phase noise compensation. *Optics Express*, 20(8):8862–8870, 2012.

[142] K. Shi, E. Sillekens, and B. C. Thomsen. 246 GHz digitally stitched coherent receiver. In *Optical Fiber Communication Conference*, pages M3D–3. Optical Society of America, 2017.

[143] A. Alvarado and E. Agrell. Four-dimensional coded modulation with bit-wise decoders for future optical communications. *Journal of Lightwave Technology*, 33(10):1993–2003, 2015.

[144] L. Szczecinski and A. Alvarado. *Bit-interleaved coded modulation: fundamentals, analysis and design*. John Wiley & Sons, 2015.

[145] B. S. Krongold, K. Ramchandran, and D. L. Jones. Computationally efficient optimal power allocation algorithms for multicarrier communication systems. *IEEE Transactions on Communications*, 48(1):23–27, 2000.

[146] W. Yu and J. M. Cioffi. On constant power water-filling. In *International Conference on Communications (ICC), 2002 IEEE*, volume 6, pages 1665–1669. IEEE, 2001.

[147] D. Lavery, R. Maher, D. Millar, A. Alvarado, S. J. Savory, and P. Bayvel. Why compensating fibre nonlinearity will never meet capacity demands. *arXiv preprint arXiv:1512.03426*, 2015.

[148] G. Kramer, M. I. Yousefi, and F. R. Kschischang. Upper bound on the capacity of a cascade of nonlinear and noisy channels. In *Information Theory Workshop (ITW), 2015 IEEE*, pages 1–4. IEEE, 2015.

[149] M. I. Yousefi, G. Kramer, and F. R. Kschischang. Upper bound on the capacity of the nonlinear Schrödinger channel. In *Information Theory (CWIT), 2015 IEEE 14th Canadian Workshop on*, pages 22–26. IEEE, 2015.

[150] S. Beppu, K. Kasai, M. Yoshida, and M. Nakazawa. 2048 QAM (66 Gbit/s) single-carrier coherent optical transmission over 150 km with a potential SE of 15.3 bit/s/Hz. *Optics Express*, 23(4):4960–4969, 2015.

[151] H. Mehmood, S. Rahman, and J. M. Cioffi. Bit loading profiles for high-speed data in DOCSIS 3.1. *IEEE Communications Magazine*, 53(3):114–120, 2015.

[152] A. D. Neira, N. Olivier, M. E. Nasir, W. Dickson, G. A. Wurtz, and A. V. Zayats. Eliminating material constraints for nonlinearity with plasmonic metamaterials. *Nature Communications*, 6:7757, 2015.

[153] A. Hasegawa and Y. Kodama. *Solitons in optical communications*. Clarendon Press, Oxford, UK, 1995.

[154] L. F. Mollenauer and J. P. Gordon. *Solitons in optical fibers: Fundamentals and applications*. Academic Press, New York, 2006.

[155] R.-J. Essiambre, R. W. Tkach, R. Ryf, I. Kaminow, T. Li, and A. E. Willner. Fiber nonlinearity and capacity: Single-mode and multimode fibers. In *Optical Fiber Telecommunications VI B*, pages 1–43. Academic Press, San Diego, CA, USA, 2013.

[156] K. S. Turitsyn, S. A. Derevyanko, I. V. Yurkevich, and S. K. Turitsyn. Information capacity of optical fiber channels with zero average dispersion. *Physical Review Letters*, 91(20):203901, 2003.

[157] D. S. Millar, S. Makovejs, C. Behrens, S. Hellerbrand, R. I. Killey, P. Bayvel, and S. J. Savory. Mitigation of fiber nonlinearity using a digital coherent receiver. *IEEE Journal of Selected Topics in Quantum Electronics*, 16(5):1217–1226, 2010.

[158] L. B. Du, M. M. Morshed, and A. J. Lowery. Fiber nonlinearity compensation for OFDM super-channels using optical phase conjugation. *Optics Express*, 20(18):19921–19927, 2012.

[159] X. Liu, A. R. Chraplyvy, P. J. Winzer, R. W. Tkach, and S. Chandrasekhar. Phase-conjugated twin waves for communication beyond the Kerr nonlinearity limit. *Nature Photonics*, 7(7):560–568, 2013.

[160] S. V. Manakov. On the theory of two-dimensional stationary self-focusing of electromagnetic waves. *Sov. Phys. - JETP*, 38(2):248–253, 1974.

[161] J. D. Ania-Castanón, T. J. Ellingham, R. Ibbotson, X. Chen, L. Zhang, and S. K. Turitsyn. Ultralong Raman fiber lasers as virtually lossless optical media. *Physical Review Letters*, 96(2):023902, 2006.

[162] T. J. Ellingham, J. D. Ania-Castañón, R. Ibbotson, X. Chen, L. Zhang, and S. K. Turitsyn. Quasi-lossless optical links for broad-band transmission and data processing. *IEEE Photonics Technology Letters*, 18(1):268–270, 2006.

[163] M. I. Yousefi and F. R. Kschischang. Information transmission using the nonlinear fourier transform, part i: Mathematical tools. *IEEE Transactions on Information Theory*, 60(7):4312–4328, 2014.

[164] S. K. Turitsyn, J. E. Prilepsky, S. Th. Le, S. Wahls, L. L. Frumin, M. Kamalian, and S. A. Derevyanko. Nonlinear Fourier transform for optical data processing and transmission: Advances and perspectives. *Optica*, 4(3):307–322, 2017.

[165] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur. The inverse scattering transform-fourier analysis for nonlinear problems. *Studies in Applied Mathematics*, 53(4):249–315, 1974.

[166] X. Yangzhang, M. I. Yousefi, A. Alvarado, D. Lavery, and P. Bayvel. Nonlinear frequency-division multiplexing in the focusing regime. In *Opt. Fiber Commun. Conf. (OFC)*, page Tu3D.1. Optical Society of America, 2017.

[167] A. Hasegawa and T. Nyu. Eigenvalue communication. *Journal of Lightwave Technology*, 11(3):395–399, 1993.

[168] S. Wahls and H. V. Poor. Fast numerical nonlinear Fourier transforms. *IEEE Transactions on Information Theory*, 61(12):6957–6974, 2015.

[169] J. E. Prilepsky, S. A. Derevyanko, K. J. Blow, I. Gabitov, and S. K. Turitsyn. Nonlinear inverse synthesis and eigenvalue division multiplexing in optical fiber channels. *Physical Review Letters*, 113(1):013901, 2014.

[170] M. I. Yousefi and F. R. Kschischang. Information transmission using the nonlinear fourier transform, part iii: Spectrum modulation. *IEEE Transactions on Information Theory*, 60(7):4346–4369, 2014.

[171] S. Hari, M. I. Yousefi, and F. R. Kschischang. Multieigenvalue communication. *Journal of Lightwave Technology*, 34(13):3110–3117, 2016.

[172] P. Kazakopoulos and A. L. Moustakas. On the soliton spectral efficiency in non-linear optical fibers. In *Information Theory (ISIT), 2016 IEEE International Symposium on*, pages 610–614. IEEE, 2016.

[173] O. V. Yushko, A. A. Redyuk, M. P. Fedoruk, and S. K. Turitsyn. Coherent soliton communication lines. *Journal of Experimental and Theoretical Physics*, 119(5):787–794, 2014.

[174] T. Gui, C. Lu, A. P. T. Lau, and P. K. A. Wai. High-order modulation on a single discrete eigenvalue for optical communications based on nonlinear Fourier transform. *Optics Express*, 25(17):20286–20297, 2017.

[175] E. Meron, M. Feder, and M. Shtaif. On the achievable communication rates of generalized soliton transmission systems. *arXiv preprint arXiv:1207.0297*, 2012.

[176] G. E. Falkovich, M. G. Stepanov, and S. K. Turitsyn. Statistics of interacting optical solitons. *Physical Review E*, 64(6):067602, 2001.

[177] Q. Zhang and T. H. Chan. Achievable rates of soliton communication systems. In *Information Theory (ISIT), 2016 IEEE International Symposium on*, pages 605–609. IEEE, 2016.

[178] Q. Zhang and T. H. Chan. A Gaussian noise model of spectral amplitudes in soliton communication systems. In *Signal Processing Advances in Wireless Communications (SPAWC), 2015 IEEE 16th International Workshop on*, pages 455–459. IEEE, 2015.

[179] S. A. Derevyanko, S. K. Turitsyn, and D. A. Yakushev. Non-gaussian statistics of an optical soliton in the presence of amplified spontaneous emission. *Optics letters*, 28(21):2097–2099, 2003.

[180] S. A. Derevyanko, S. K. Turitsyn, and D. A. Yakushev. Fokker-planck equation approach to the description of soliton statistics in optical fiber transmission systems. *JOSA B*, 22(4):743–752, 2005.

[181] S. A. Derevyanko, J. E. Prilepsky, and S. K. Turitsyn. Capacity estimates for optical transmission based on the nonlinear Fourier transform. *Nature Communications*, 7:12710, 2016.

[182] M. I. Yousefi and X. Yangzhang. Linear and nonlinear frequency-division multiplexing. In *ECOC 2016; 42nd European Conference on Optical Communication; Proceedings of*, pages 1–3. VDE, 2016.

[183] I. Tavakkolnia and M. Safari. Capacity analysis of signaling on the continuous spectrum of nonlinear optical fibers. *Journal of Lightwave Technology*, 35(11):2086–2097, 2017.

[184] M. Kamalian, J. E. Prilepsky, S. T. Le, and S. K. Turitsyn. Spectral efficiency estimation in periodic nonlinear Fourier transform based communication systems. In *Optical Fiber Communications Conference and Exhibition (OFC), 2017*, pages 1–3. IEEE, 2017.

[185] S. A. Derevyanko, J. E. Prilepsky, and D. A. Yakushev. Statistics of a noise-driven Manakov soliton. *Journal of Physics A: Mathematical and General*, 39(6):1297, 2006.

[186] T. I. Lakoba and D. J. Kaup. Perturbation theory for the Manakov soliton and its applications to pulse propagation in randomly birefringent fibers. *Physical Review E*, 56(5):6147, 1997.

[187] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura. Method for solving the Korteweg-deVries equation. *Physical Review Letters*, 19(19):1095, 1967.

[188] P. D. Lax. Integrals of nonlinear equations of evolution and solitary waves. *Communications on pure and applied mathematics*, 21(5):467–490, 1968.

[189] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur. The inverse scattering transform-fourier analysis for nonlinear problems. *Studies in Applied Mathematics*, 53(4):249–315, 1974.

[190] J. E. Prilepsky, S. A. Derevyanko, and S. K. Turitsyn. Nonlinear spectral management: Linearization of the lossless fiber channel. *Optics express*, 21(20):24344–24367, 2013.

[191] A. S. Fokas and I. M. Gelfand. Integrability of linear and nonlinear evolution equations and the associated nonlinear Fourier transforms. *Letters in Mathematical Physics*, 32(3):189–210, 1994.

[192] A. R. Osborne. The inverse scattering transform: tools for the nonlinear Fourier analysis and filtering of ocean surface waves. *Chaos, Solitons & Fractals*, 5(12):2623–2637, 1995.

[193] M. J. Ablowitz, B. Prinari, and A. D. Trubatch. Integrable nonlinear Schrödinger systems and their soliton dynamics. *Dynamics of Partial Differential Equations*, 1(3):239–299, 2004.

[194] A. Hasegawa and M. Matsumoto. Optical solitons in fibers. In *Optical Solitons in Fibers*, pages 41–59. Springer, 2003.

[195] A. Maruta. Eigenvalue modulated optical transmission system. In *Opto-Electronics and Communications Conference (OECC), 2015*, pages 1–3. IEEE, 2015.

[196] L. Schmalen, A. Alvarado, and R. Rios-Müller. Performance prediction of nonbinary forward error correction in optical transmission experiments. *Journal of Lightwave Technology*, 35(4):1015–1027, 2017.

[197] R.-J. Essiambre, G. J. Foschini, G. Kramer, and P. J. Winzer. Capacity limits of information transport in fiber-optic networks. *Physical Review Letters*, 101(16):163901, 2008.

[198] N. L. Johnson, S Kotz, and Balakrishnan N. *Continuous Univariate Distributions: Volume 2*, volume 2. John Wiley & Sons, New York, 2006.

[199] M. Katz and S. Shamai. On the capacity-achieving distribution of the discrete-time noncoherent and partially coherent awgn channels. *IEEE Transactions on Information Theory*, 50(10):2257–2270, 2004.

[200] A. Lapidoth. Capacity bounds via duality: A phase noise example. In *Proc. 2nd Asian-European Workshop on Information Theory*, pages 58–61. Citeseer, 2002.

[201] A. Lapidoth and S. M. Moser. Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels. *IEEE Transactions on Information Theory*, 49(10):2426–2467, 2003.

[202] A. Mecozzi and M. Shtaif. On the capacity of intensity modulated systems using optical amplifiers. *IEEE Photonics Technology Letters*, 13(9):1029–1031, 2001.

[203] M. I. Yousefi and F. R. Kschischang. Information transmission using the nonlinear fourier transform, part ii: Numerical methods. *IEEE Transactions on Information Theory*, 60(7):4329–4345, 2014.

[204] J. P. Boyd. *Chebyshev and Fourier spectral methods*. Courier Corporation, 2001.

[205] Y. Yu. The shape of the noncentral chi-square density. *arXiv preprint arXiv:1106.5241*, 2011.

[206] A. Marsiglietti and V. Kostina. A lower bound on the differential entropy for log-concave random variables with applications to rate-distortion theory. In *IEEE International Symposium on Information Theory (ISIT)*, pages 46–50. IEEE, 2017.

[207] J. E. Prilepsky and S. A. Derevyanko. Statistics of noise-driven coupled nonlinear oscillators: Applications to systems with kerr nonlinearity. *Physica D: Nonlinear Phenomena*, 203(3-4):249–269, 2005.

[208] D. J. Kaup. Perturbation theory for solitons in optical fibers. *Physical Review A*, 42(9):5689, 1990.

[209] H. A Haus and Y. Lai. Quantum theory of soliton squeezing: a linearized approach. *JOSA B*, 7(3):386–392, 1990.

[210] G. H. Golub and C. F. van Loan. *Matrix computations*. 1996.

[211] W. E. Arnoldi. The principle of minimized iterations in the solution of the matrix eigenvalue problem. *Quarterly of Applied Mathematics*, 9(1):17–29, 1951.

[212] D. C. Sorensen. Implicit application of polynomial filters in a k-step Arnoldi method. *SIAM Journal on Matrix Analysis and Applications*, 13(1):357–385, 1992.

[213] D. B. de Haan. *Nouvelles tables: d'intégrales définies*. Leide, 1867.

[214] M. Geller and E. W. Ng. A table of integrals of the exponential integral. *Journal of Research of the National Bureau of Standards*, 73(3):191–210, 1969.